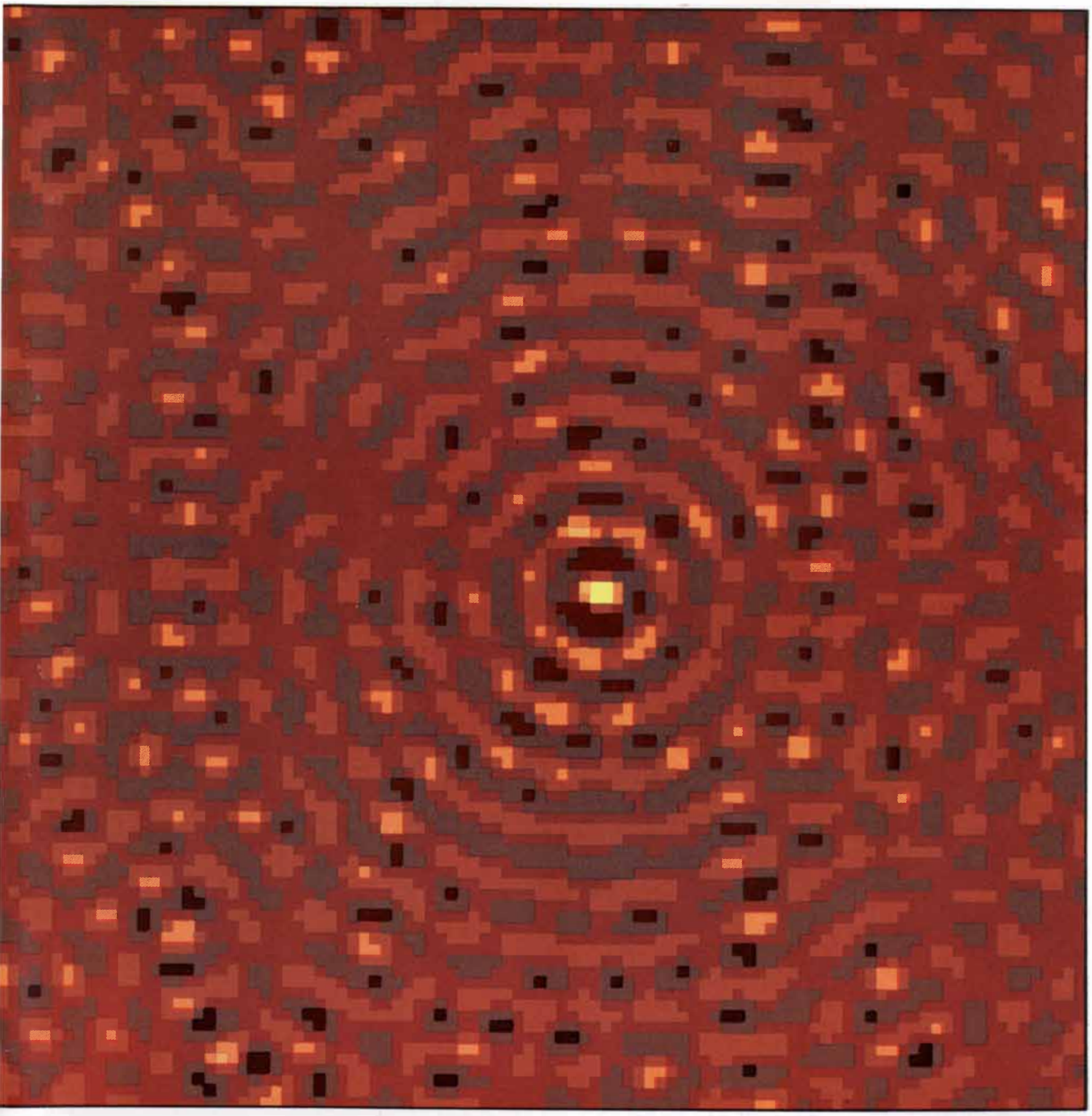
SCIENTIFIC AMERICAN



THE X-RAY SKY

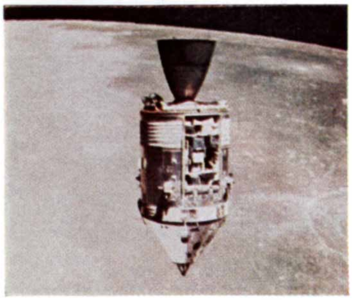
ONE DOLLAR

July 1972

Being plush isn't enough. I think a car should also be tough. That's why I like the Chrysler New Yorker.

– ARTHUR GODFREY

If you've ever ridden in a Chrysler New Yorker Brougham, you know what I mean when I say it's plush. But just as important is the way this car is built. I've seen Chrysler build their cars, I think they're really trying to put together a car that will work better and last longer than any they've ever built before. That's important.



Here's another famous way to get around. Apollo 15. I wanted you to see it because it has to be strong. The structure of the command module is welded into a single unit. Chrysler welds their car bodies into a single unit for the same reason. Strength. In fact, Chryslers are welded in places where other cars are bolted together. The Chrysler engineers believe this makes for a strong car. And I think we all deserve cars that are as strong as possible.

This is what the inside of the New Yorker Brougham looks like. The seats have a beautiful fabric with a Jacquard weave. But more important, that fabric is built to look good through years of use. Everything that goes into this car has to be two things. It has to be high quality. And it has to be built to last.

You can tell a lot about the New Yorker just from looking at it. It's big. It's comfortable. And I think it's good looking. It's also a very quiet car. That's one of the things you can't see. It has rubber body cushions that help isolate the noise and vibration of the road, helping to keep it out of the interior of the car.

Try a New Yorker. I think it'll spoil you for any other kind of car. It's plush, and it's built to last. That's where Chrysler got their slogan for this year.

Coming through with the kind of car America wants.

CHRYSLER

CHRYSLER

Plymouth

IF WE GOT A CALL TODAY AT 9AM TO BUILD A FLOATING CITY, WE COULD GO TO WORK ON IT AT 9:05.

Nowhere is the problem of population growth more evident than in coastal cities.

Tokyo. New York. San Francisco. Los Angeles. And Hong Kong. And these cities will run out of land. But there is a solution.

Floating cities.

Totally contained environments to relieve urban congestion.

And we could go to work to build them today.

Because we've literally built floating cities for more than a generation. Aircraft carriers. We've built 20.

And the two nuclear-powered carriers now under construction will each have a crew of about 6,000.

So you see, we have the expertise.

Because we're already at work on marine systems, enclosed environments, energy production and industrial construction.

Perhaps the first floating city would only have a population of 6,000. But ultimately, floating cities could be equal to Chicagos on the seas.

Exciting? Yes. Out of sight? No.

Floating cities are just around the corner. And when the time comes to build them, we're ready.

Newport News Shipbuilding TENNECO **Where tomorrow is. Now.**

A Tenneco Company Newport News, Virginia 23607

B4700 Systems . . . B3700 Systems . . . B2700 Systems . . . MCP V

Burroughs announces three powerful computer series and a fifth major advance in system software

Important news for every user of medium-scale computer systems. Burroughs 700 Systems family is expanding, with —

More powerful and productive computer systems

The B4700 Series: Five additional models with up to 100% more input/output potential than current models . . .

The B3700 Series: Two new models based on the B4700's proven design . . . with up to 70% more data processing power than Burroughs widely used B3500 systems . . .

The B2700 Series: Six models that extend B2500/B3500 series advanced concepts to an even broader price-and-performance range.

These outstanding systems offer you —

- "Instant maturity" — they accept B2500/B3500 application programs, developed and perfected over six years, and execute them at optimum performance levels.
- Economical, modular expansion within a series, and from one series to another, without reprograming or even recompilation.
- A comprehensive library of program products for major savings in program development time and related costs.
- Powerful FORTRAN, COBOL and BASIC compilers . . . plus advanced software for remote job entry, data communications and data base management.
- Automatic multiprograming under MCP control . . . system self-regulation . . . virtual memory . . . dynamic allocation of resources . . . and other results-oriented features developed and proven by Burroughs.
- Parallel multiprograming and shared disk file processing.
- Data communications processors for greater efficiency and line-handling capacity.
- Powerful terminal systems for application in business, industry, finance and government.

A more powerful and productive master control program

Tests indicate that MCP V increases throughput performance 10 to 20% both for new systems and currently installed systems. MCP V is now available for *every* Burroughs medium system. Based on more than six years' successful field experience, MCP V confirms Burroughs leadership in control program technology.

Burroughs puts you ahead with cost/effective systems that do more and cost less to install, program and use . . . keeps you ahead with systems that keep on doing more and costing less as they expand to meet growth and change in your data processing operations.

Established 1845 **SCIENTIFIC AMERICAN** July 1972 Volume 227 Number 1

ARTICLES

14 **ANTISUBMARINE WARFARE AND NATIONAL SECURITY, by Richard L. Garwin** A defense against missile submarines would work against deterrence.

26 **THE X-RAY SKY, by Herbert W. Schnopper and John P. Delvaille** Rockets and a satellite have now detected 120 X-ray-emitting celestial objects.

38 **LEWIS CARROLL'S LOST BOOK ON LOGIC, by W. W. Bartley III** The author of *Alice* also wrote *Symbolic Logic,* to which a sequel has been found.

52 **THE CHEMICAL ELEMENTS OF LIFE, by Earl Frieden** Fluorine, silicon, tin and vanadium have recently enlarged the list from 20 to 24.

65 **THE TOKAMAK APPROACH IN FUSION RESEARCH, by Bruno Coppi and Jan Rem** Several large machines will explore the "toroidal diffuse pinch."

76 **DEPRIVATION DWARFISM, by Lytt I. Gardner** Children raised in an emotionally deprived environment can be stunted in growth.

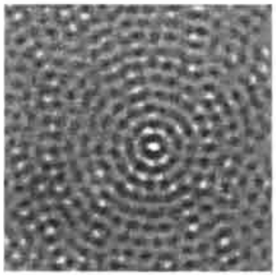
84 **EXPERIMENTS IN READING, by Paul A. Kolers** They suggest that reading is not linear but calls for the framing of hypotheses.

92 **ESCAPE RESPONSES IN MARINE INVERTEBRATES, by Howard A. Feder** Limpets, scallops and other creatures are remarkably lively in eluding starfishes.

DEPARTMENTS

6 LETTERS
8 50 AND 100 YEARS AGO
13 THE AUTHORS
48 SCIENCE AND THE CITIZEN
102 MATHEMATICAL GAMES
106 THE AMATEUR SCIENTIST
117 BOOKS
122 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: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.



THE COVER

The pattern on the cover is a computer-generated display of the location of a strong celestial X-ray source designated GX9 +1. The display, known as a correlation map, is based on data gathered by means of an X-ray detection device aboard a rocket launched by a group at the Massachusetts Institute of Technology. The source itself is at the bright yellow square near the center, surrounded by darker circular rings. The pattern is composed of 10,000 tiny squares or test points dividing up the field of view; the probability that the source is at each of these test points is calculated by computer, based on the data from the detection device. A color is assigned to each probability; lighter colors are areas of higher probability and darker colors areas of lower probability. An explanation of the experimental procedure used to compile correlation maps and an account of the results of recent studies of X-ray sources are given by Herbert W. Schnopper and John P. Delvaille in their article "The X-Ray Sky" (*page 26*).

THE ILLUSTRATIONS

Cover illustration by Gabor Kiss

Page	Source
15–23	Gabor Kiss
24–25	U.S. Navy
26	Herbert W. Schnopper and John P. Delvaille, Massachusetts Institute of Technology
28–31	Enid Kotschnig
32–34	Herbert W. Schnopper and John P. Delvaille, Massachusetts Institute of Technology
35–36	Enid Kotschnig
38	Gernsheim Collection, Humanities Research Center, University of Texas at Austin
40	Christ Church, University of Oxford
46	The Charles Lutwidge Dodgson Estate and W. W. Bartley III
52–55	Graphic Presentation Services, Inc.
56	Ilil Arbel
57	Graphic Presentation Services, Inc.
59–60	Veterans Administration Hospital, Long Beach, Calif.
65	Allen Beechel
66	I. V. Kurchatov Institute of Atomic Energy
67	Allen Beechel
68	Allen Beechel (*top*), Ilil Arbel (*bottom*)
69	Allen Beechel
70–71	Ilil Arbel
72–75	Allen Beechel
77–82	Bunji Tagawa
85	Jim Egleson
86	Jim Egleson (*top*); Paul A. Kolers, University of Toronto (*bottom*)
87	Paul A. Kolers, University of Toronto
88	Jim Egleson
89	Paul A. Kolers, University of Toronto
90	Ilil Arbel
91	Jim Egleson
92–94	Holger Knudsen, Marine Biological Laboratory, Helsingör, Denmark
96	D. L. Ray, from Abe Margolin (*top*), Tom Prentiss (*bottom*)
97	Tom Prentiss
98	D. M. Ross, University of Alberta
99–100	Tom Prentiss
103–105	Ilil Arbel
106–110	Roger Hayward

LETTERS

Sirs:

In the article "Tides and the Earth-Moon System" [SCIENTIFIC AMERICAN, April] Peter Goldreich rejects the fission hypothesis for the origin of the moon in part on the ground that "computations show that an initially equatorial orbit would have evolved into an orbit that lay in the ecliptic plane at the present earth-moon separation." This statement is true if we assume that the tidal-lag angle, which is the angle between the crest of the tidal wave and the direction to the moon, has always been only a few degrees, as it is now.

But Sir George Darwin, to whom Professor Goldreich refers, showed (*Scientific Papers*, Vol. 2, Paper No. 6, page 319) that if the earth behaved like a viscous fluid for a few centuries after it was formed, then the effect of the tides would have been to wrench it out of the earth's equatorial plane.

The mathematics is reasonably straightforward; the paper has never been shown to be wrong, and I think the reasonable presumption is that it is right. In that case the dynamical objection raised by Professor Goldreich is seen to depend on the model assumed for the viscosity of the early earth.

JOHN A. O'KEEFE

Goddard Space Flight Center
Greenbelt, Md.

Sirs:

By far the strongest dynamical argument against the fission hypothesis of lunar origin is the angular-momentum difficulty (mentioned in the first paragraph on page 51 of my article). In my opinion no plausible mechanism has ever been proposed to account for the loss of half of the initial angular momentum in the earth-moon system following the fission event. However, should fission actually have taken place, then it is true, as Dr. O'Keefe points out, that the subsequent evolution of the lunar orbit would have depended on the viscoelastic properties of the earth. In particular, the inclination of the moon's orbit to the earth's Equator might have increased with time if the physical properties of the earth were very different then from what they are now.

PETER GOLDREICH

California Institute of Technology
Pasadena

Sirs:

Björn Kurtén is to be congratulated for his excellent contribution on the evolution of the cave bears during the Pleistocene [SCIENTIFIC AMERICAN, March]. As an archaeologist I find Kurtén's critical discussion and dismissal of upper Paleolithic "bear cults" interesting, but readers may not be aware of earlier evidence that is clearly associated with middle Paleolithic cultural remains. The site of Régourdou cave, located near the famous painted cave of Lascaux in the Dordogne of France, has yielded impressive evidence for ritual use of bear bones by the Neanderthalers (E. Bonifay, "Un ensemble rituel Moustérien à la grotte du Régourdou," *International Congress of Prehistoric and Protohistoric Science*, Sixth Proceedings, Rome, 1962, pages 132–140). At Régourdou the roof of the cave collapsed during the Würm glacial complex, sealing the archaeological levels under a rockfall. Excavations yielded a Neanderthal burial and evidence for Mousterian tools belonging to the La Quina cultural variant. Thus the antiquity and general setting of the site cannot really be questioned in the same way as the Drachenloch example cited by Kurtén. Bonifay's excavations have revealed multiple instances where bear bones have been specially treated by these early inhabitants of the Dordogne region. Bear bones were recovered from at least three different levels where they were associated with structural remains that had been intentionally placed in the cave by early man. This evidence suggests that there was a regular custom of placing bear bones in the cave that spanned an extended period of time. The bear bones were recovered from pits and from within various stone tumuli. In one instance bear skeletal remains were recovered from a pit that was bordered by a dry stone wall and covered by a huge stone slab. Evidently the bear had been butchered prior to having been placed in the pit. This level also yielded a rock structure containing bear bones and the Neanderthal burial that contained a number of offerings including the humerus of a bear.

Kurtén concludes that possible bear cults were not a major factor in the extinction of cave bears. In this regard it is of some interest to note that the Régourdou bear burials were not cave bears (*Ursus spelaeus*) but brown bears (*U. arctos* L.). In this regard, if bears were being killed by Paleolithic hunters for ritual purposes, it would appear that the major selective force would be against the brown bears. Yet they are still very much evident today.

LAWRENCE H. ROBBINS

Michigan State University
East Lansing

Sirs:

I wish to thank Dr. Robbins for his interesting comment, with which I am in complete agreement. As I noted in my article, it seems to have been the brown bear (*Ursus arctos*) rather than the cave bear (*U. spelaeus*) that was preyed on by man. I have recently finished a study of the great accumulation of bear bones in the kitchen midden of Taubach near Weimar in Germany, which dates from the Eemian interglacial period (Neanderthalian times!). This is all brown bear, exactly as at Régourdou.

BJÖRN KURTÉN

Museum of Zoology
Helsinki

Scientific American, July, 1972; Vol. 227, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer.

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 25 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 ☐
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1922: "By means of a special apparatus that reproduces actual lightning and is known as a lightning generator, the laboratory of Dr. Charles P. Steinmetz, chief consulting engineer of the General Electric Company, is making the most successful tests of lightning arresters that have ever been conducted. Except in the matter of magnitude the artificial lightning is exactly the same kind of energy, stored up and discharged in the same way, as the gigantic lightning bolt of an electrical storm in the heavens. Any object that is placed in the path of the artificial lightning stroke is torn to pieces just as truly as it would be if it were in the path of natural lightning. Unusually severe service conditions can be produced, and the action of the lightning arresters can be observed with accuracy. The artificial lightning represents, while it lasts, more than a million horsepower. The estimated horsepower of a natural lightning flash is given as 500 million horsepower. It lasts about a hundred-thousandth of a second."

"San Francisco is engaged in the construction of a municipal water supply system, known as the Hetch Hetchy Water Supply project, of greater magnitude than any similar system in the United States, with the exception of those of New York and Boston. It will make available to San Francisco and the adjacent metropolitan area 400,000,000 gallons daily of pure mountain water, and this same water, passing through huge water wheels installed in electric power plants, will generate 200,000 horsepower for general use in central California. The Tuolumne River and some of its tributaries in the Sierra Nevada, more than 150 miles from San Francisco, are the sources of the city's future water supply. The project takes its name from the principal reservoir site, Hetch Hetchy Valley, within the boundaries of Yosemite National Park."

"Recent investigations in the field of proteins and nutrition conducted by the Bureau of Chemistry of the United States Department of Agriculture have shown that different proteins vary widely in their nutritive value. A diet may furnish a sufficient amount of protein, fat, carbohydrates, salts and vitamines and yet fail to promote growth or sustain well-being unless the quality of protein is nutritionally adequate. Certain of the amino acids, of which about 19 have been found in proteins, are absolutely essential for growth and maintenance, among which are lysine, cystine and tryptophane. A protein that is deficient in lysine or cystine, even though it contains all of the other amino acids, will fail from a nutritive standpoint. The deficiency, however, when properly understood, can be supplemented by the addition of other foods to the diet that contain adequate amounts of these essential amino acids."

SCIENTIFIC AMERICAN

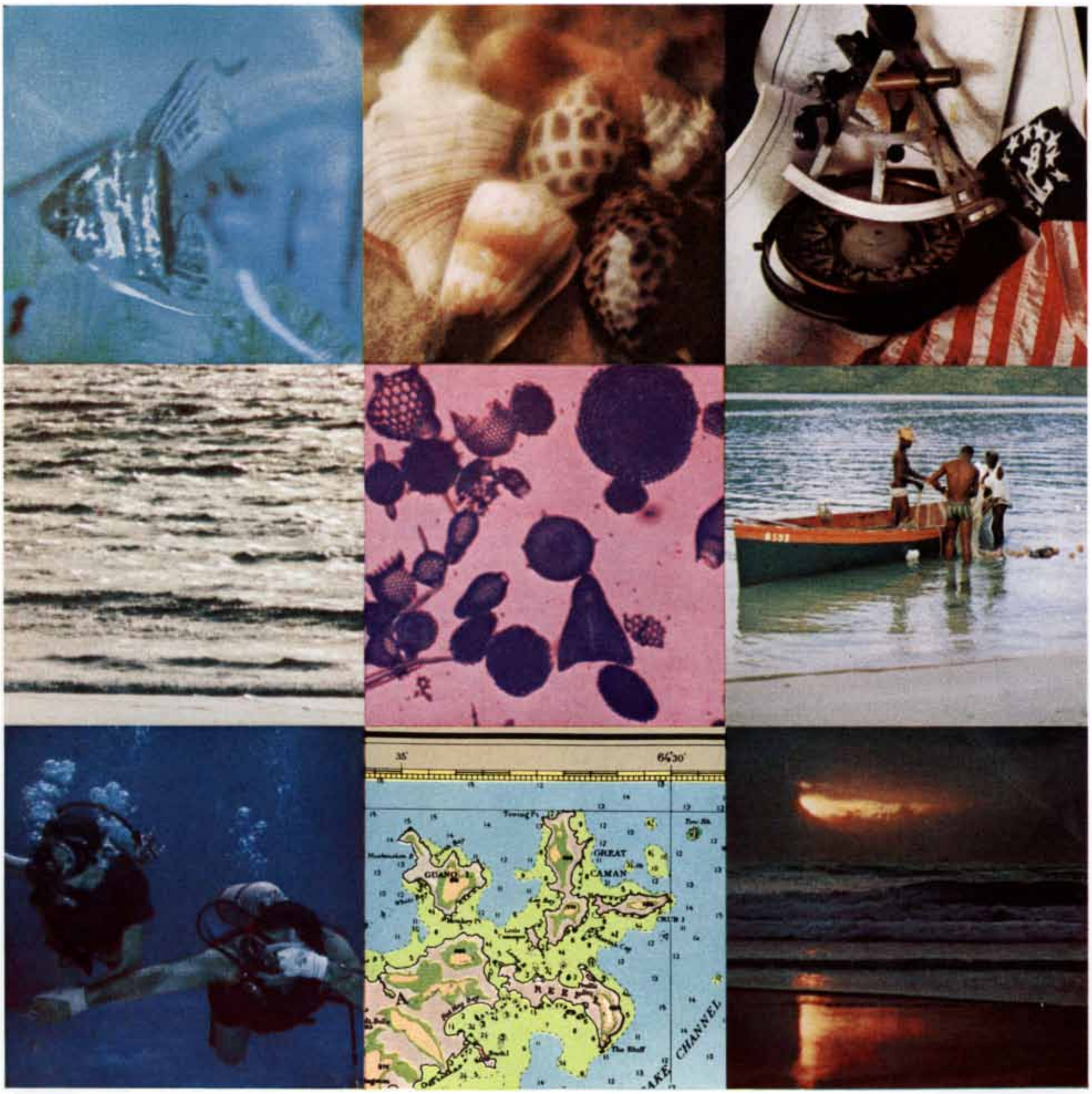
JULY, 1872: "The great uprising for the eight-hour workday that has been agitating the city of New York for the past 10 weeks has at length reached its termination. Toward its close it was gradually but surely weakening. One by one the different trade organizations withdrew their support and returned to work under the old system until, with the exception of the iron and metal workers, none of the great leagues, which in the beginning made such a show of strength and placed such confidence in their own powers, were left in existence. It was reserved for the abovementioned association to give the finishing stroke to the movement, which it did by declaring by vote in solemn conclave that the strike was at an end. At the same time this colossal association, which but lately threatened, through the aid of English gold and its emissaries sent throughout the country, to revolutionize the relations of capital and labor, decided to disband, leaving its members to follow their own inclination in the matter of returning to work at the old hours and former wages."

"As a means of disseminating useful information through the medium of cheap literature to the masses, as an aid in promoting social intercourse, as a facilitation of business enterprise, as a help to self-education through increased letter writing and as affording fuller and freer interchange of ideas, our postal system is of the utmost national importance. The acquisition of such reforms as will ensure its greatest efficiency at the lowest possible cost is a subject interesting to every individual able to read and write. The new postal rates, which have lately come into effect, are more valuable on account of their opening the way toward future and greater reductions than for the saving of postage in which they may at present result. A thousand million letters yearly pass through our mails, and yet the statistics of the Post Office Department show that the country suffers a deficiency, and that, instead of being a means of revenue, our postal arrangements are a source of expense. This, by proper organization, judicious retrenchment and a uniform low rate of postage honestly enforced, can be eventually remedied."

"At the recent session of the New York state legislature several projects for securing rapid transit in New York City were passed. The most prominent is the charter granted to the millionaire Vanderbilt, giving him authority to construct an underground steam railway commencing on Fourth Avenue at 59th Street and extending southward to City Hall Park at a point near Broadway between the City Hall and the Post Office. The expense of construction will be very heavy, but it is stated that the railway will be soon commenced and built."

"According to the returns of the census up to June 1, 1870, the United States has 12,505,923 working people. The number of inhabitants in the country is 38,558,371, so that the active workers constitute very nearly one third of the entire population, the ratio having considerably increased since the census of 1860, at which time it barely exceeded one quarter. Of the working people 10,669,436 are males and 1,836,487 females. These figures apply to the men and women in actual outside employment. The population of the country may be estimated to be divided into 8,000,000 families, each of which has a woman for one of its heads. She is not considered as a worker in the foregoing calculations as given by the census, and here we consider a mistake has been made. The poor man's wife has far more labor to perform than her unmarried sister who works her 10 hours a day. In at least 7,000,000 families the lot of the wife and mother is no sinecure; so that in reality we find not only that the working women and the working men are equal in number but also that there is a balance on the side of the women in the shape of unending labor, the most monotonous and thankless in existence."

How we look at it.



We look at the same thing most oil companies look at, but we never get used to it.

At American Oil, we think there's still a lot to learn about a common drop of crude oil. That's why we look at it so closely. Because the more we can see in a single drop . . . the better the oils and lubricants we can keep offering your industry.

At American Oil, it's what we keep learning that makes us important to you.

AMERICAN

American Oil Company


The phone company wants more operators like Rick Wehmhoefer.
Rick Wehmhoefer of Denver, Colorado, is one of several hundred male telephone operators in the Bell System.
Currently, Rick is a directory assistance operator. “So far, my job has been pleasant and worthwhile,” he says. “I enjoy assisting people.”
We have men like Rick in a lot of different telephone jobs. Both men and women work as Bell System mechanics, truck drivers, installers and engineers.
We want the men and women of the telephone company to do what they want to do, and do best.
Today, when openings exist, local Bell Companies are offering applicants and present employees some jobs they may never have thought about before. We want to help all advance to the best of their abilities.
AT&T and your local Bell Company are equal opportunity employers.

THE AUTHORS

RICHARD L. GARWIN ("Antisubmarine Warfare and National Security") is with the International Business Machines Corporation as an IBM Fellow, holding broad responsibility for contributions to technology and to society. He is also a member of the President's Science Advisory Committee and a consultant to a number of Government and non-Government organizations concerned with defense technology. Garwin received his degrees in physics: his bachelor's degree at the Case Institute of Technology in 1947 and his master's and Ph.D. from the University of Chicago, where he taught physics for three years until he joined IBM in 1952. Before becoming an IBM Fellow in 1967 he served for a year as director of applied research at the company's Thomas J. Watson Research Center and a year as director of the IBM Watson Laboratory at Columbia University (where he has been adjunct professor of physics since 1957).

HERBERT W. SCHNOPPER and JOHN P. DELVAILLE ("The X-Ray Sky") are at the Massachusetts Institute of Technology; Schnopper is associate professor of physics and Delvaille is a research member of the Center for Space Research. Schnopper was graduated from Rensselaer Polytechnic Institute in 1954, receiving his master's degree and his Ph.D. from Cornell University in 1958 and 1962 respectively. He taught at Cornell until he went to M.I.T. in 1966. "For relaxation," he writes, "I enjoy cross-country skiing in the winter and hiking and bicycling the rest of the year. Geometry has always fascinated me, and one of my hobbies is the wire sculpture of all sorts of geometrical forms. My wife is completing her doctoral studies in Slavic literature at Harvard, and we sometimes enjoy reading the same works—she in the native language." Delvaille was graduated from the University of California at Berkeley in 1954 and obtained his Ph.D. in physics from Cornell in 1962. He taught at Cornell until he went to M.I.T. in 1970. His avocations include backpacking, skiing, bowling and acting.

W. W. BARTLEY III ("Lewis Carroll's Lost Book on Logic") is professor of philosophy and of the history of philosophy and science at the University of Pittsburgh, where in addition he serves as associate director of the Center for the Philosophy of Science. He is also professor of philosophy at California State University at Hayward. Bartley was graduated from Harvard College in 1956 and received his doctorate in logic and scientific method at the London School of Economics in 1962. Before taking up his present work he taught at the University of London, the University of Cambridge and the University of California. In the coming academic year he will be in China on a fellowship from the American Council of Learned Societies.

EARL FRIEDEN ("The Chemical Elements of Life") is professor of chemistry at Florida State University. He took his bachelor's degree in chemistry at the University of California at Los Angeles and his Ph.D. in biochemistry at the University of Southern California, thereafter holding postdoctoral research positions at the University of Wisconsin, the Carlsberg Laboratories in Copenhagen and the University of California at San Diego. His article reflects a long-term interest in the relation between metals and biology. "In our concern over pollution," he writes, "we should never overlook the fact that some elements, particularly the metals, are 'good,' that is, essential to life in trace quantities, although they may be harmful to the animal in much larger amounts."

BRUNO COPPI and JAN REM ("The Tokamak Approach in Fusion Research") are respectively professor of physics at the Massachusetts Institute of Technology and a member of the theoretical group at the Institute for Plasma Physics at Jutphaas in the Netherlands. Coppi, who was born in Italy, did his undergraduate work at the University of Pavia and received his doctorate at the Polytechnic of Milan. In 1961 he came to the U.S. to work in theoretical plasma physics at Princeton University. When he went to M.I.T. in 1969, he held a joint appointment with Princeton and the Institute for Advanced Study. Rem, who began his higher education at the Technological University of Delft in the Netherlands, writes: "I would have become an aeronautical engineer if I had not been offered the opportunity after my fourth year to go to Northwestern University. There I was supposed to study gas flow with combustion, but the subject was not taught any more in the department of mechanical engineering, so I had to go in the direction the department was going: plasma physics." Rem obtained his Ph.D. from Northwestern in 1963 and then joined the Institute for Plasma Physics. Until recently he was a member of a team of Institute for Plasma Physics scientists working at M.I.T.

LYTT I. GARDNER ("Deprivation Dwarfism") is professor of pediatrics at the Upstate Medical Center of the State University of New York. He writes that as an undergraduate at the University of North Carolina he led "somewhat of a double life," since he was studying zoology but had originally been interested in the humanities and writing and served on the editorial staff of the student newspaper and the literary magazine. "I have written some poetry, mostly nonserious, that has found its way into medical journals as space-filler." Gardner obtained his M.D. at the Harvard Medical School in 1943 and has been with the Upstate Medical Center since 1952. "During the past 15 years," he writes, "I have been involved in considerable postgraduate teaching in Latin America, particularly in Brazil."

PAUL A. KOLERS ("Experiments in Reading") is professor in the department of psychology at the University of Toronto, where he went in 1970 after several years at the Massachusetts Institute of Technology and two years at the Bell Telephone Laboratories. He writes: "At Bell Labs I worked on problems in perception; the results are described in a book due out this year, called *Aspects of Motion Perception.* My training had been largely in the psychology of perception, with a secondary interest in the psychology of language (or 'psycholinguistics'). These two came together in the work I now concentrate on, which is how people (or their visual systems) work to acquire information from symbols such as words, pictures and diagrams." Kolers took his Ph.D. in experimental psychology at New York University in 1957.

HOWARD M. FEDER ("Escape Responses in Marine Invertebrates") is associate professor of zoology and marine science at the University of Alaska. He writes that his interest in marine biology began in the late 1940's, when as an undergraduate at the University of California at Los Angeles he became a collector of marine specimens for the department of zoology. Feder obtained his bachelor's and master's degrees in zoology at U.C.L.A., moving to Stanford University for his Ph.D., which he received in 1956.

Established 1845 SCIENTIFIC AMERICAN July 1972 Volume 227 Number 1

Antisubmarine Warfare and National Security

The missile-submarine deterrents of the U.S. and the U.S.S.R. are not threatened by current antisubmarine technology. It is argued that this state of affairs should be maintained by the two powers

by Richard L. Garwin

The treaty and executive agreement recently signed by the U.S. and the U.S.S.R., the culmination of three years of the strategic-arms-limitation talks (SALT), formally recognize that the two countries accept the military and political value of mutual deterrence as a means of forestalling an all-out nuclear war. The accords remove one potential threat to the effectiveness of strategic offensive missiles, on which deterrence rests, by prohibiting the U.S. and the U.S.S.R. from constructing nationwide anti-ballistic-missile (ABM) defenses. By ensuring the effectiveness of our offensive missiles, the ABM treaty greatly enhances our future security. The SALT accords, however, do not prohibit either country from qualitatively improving its offensive missile forces, for example by equipping missiles with multiple independently targeted reentry vehicles (MIRV's) or by improving the accuracy or explosive yield of missile warheads, perhaps to the point where the offensive missiles have a high probability of being able to destroy land-based missiles in their silos (if they were to remain there until they were hit). The U.S. could in fact replace Minuteman with a larger missile in the same silo, thereby multiplying the "throw weight" of the Minuteman force by a factor of three or four. As a result of this possibility each country will be concerned to maintain the invulnerability of its submarine-based strategic missiles, which are essentially immune to attack from land-based weapons [see "Missile Submarines and National Security," by Herbert Scoville, Jr.; SCIENTIFIC AMERICAN, June].

A potential threat to the submarine-based deterrent already exists in the antisubmarine-warfare (ASW) forces that both the U.S. and the U.S.S.R. possess. Fortunately, in the present state of the art, the technology of antisubmarine warfare is even less effective in a disabling strike against ballistic-missile-launching submarines than it is against nuclear-powered submarines designed to attack surface ships. Nevertheless, given the importance of the submarine-based deterrent, it is not too early to examine the feasibility of arms-control limitations on possible methods of antisubmarine warfare that might eventually threaten the survivability of missile-launching submarines.

In this article I shall distinguish between antisubmarine-warfare techniques whose utility is tactical (the protection of surface shipping) and techniques that would be specifically useful against submarines carrying strategic offensive missiles. I hope to show that one could limit antisubmarine-warfare methods that would threaten the submarine-based deterrent without impairing one's ability to protect surface shipping from submarine attack. The precedent set by the prohibition of nationwide ABM systems shows that both the U.S. and the U.S.S.R. have a strong interest in preserving a controllable, survivable force of submarine-launched ballistic missiles (SLBM's). Even if a conventional war were to break out between the two powers, one could hope to continue to deter mutual nuclear annihilation by arranging for the SLBM forces of the two countries to survive. To that end one might now consider a possible SALT II agreement to provide safe passage for missile-launching submarines along designated routes to the open ocean. This is only one example of what might be in the interest of both the U.S. and the U.S.S.R.

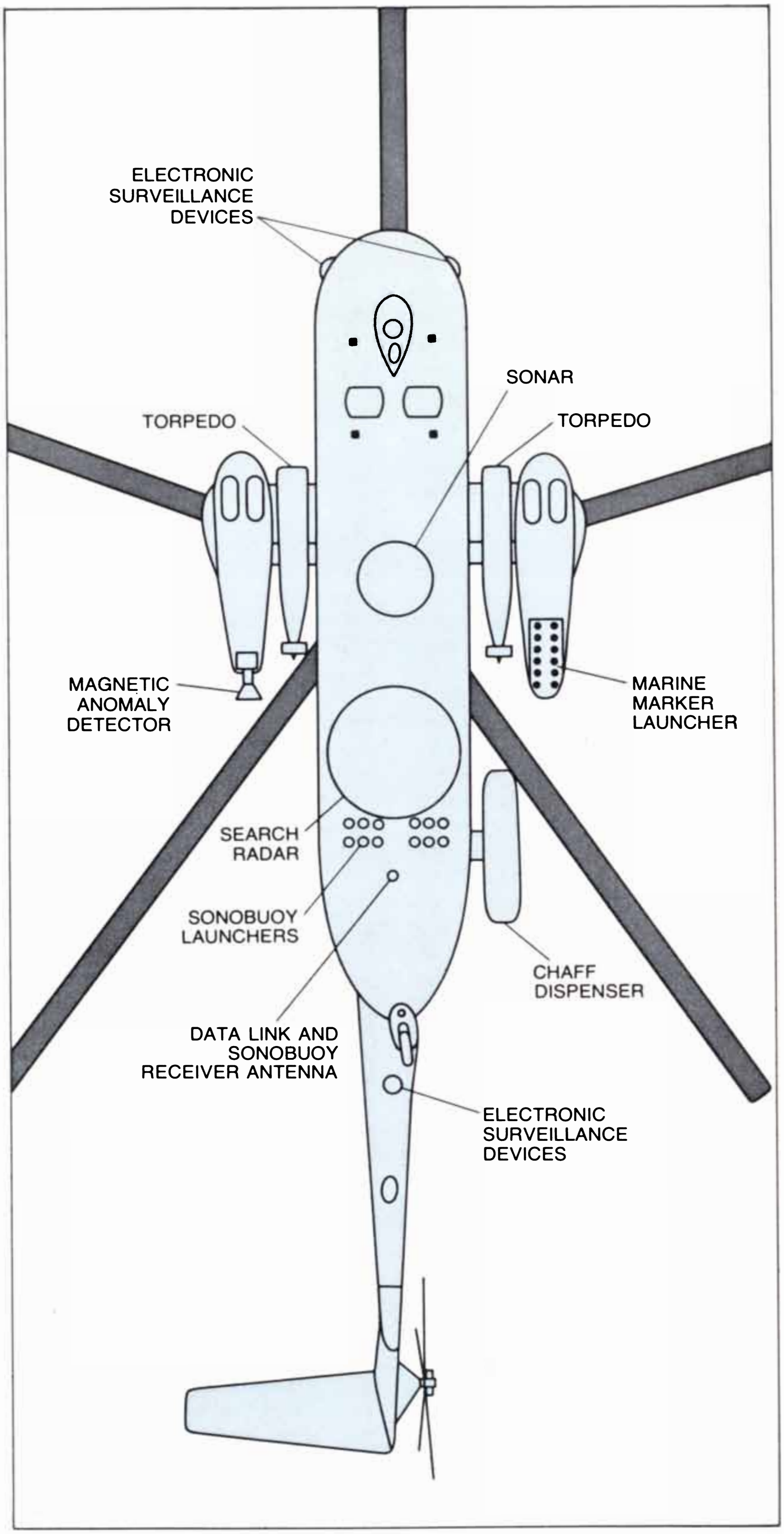
Mutual acceptance of invulnerability for ballistic-missile submarines of the U.S. and the U.S.S.R. does not extend, however, to all possible roles for those submarines. For instance, a deterrent role for the ballistic-missile submarine does not require that all boats be in firing position near the surface and able to coordinate their fire within seconds. Nor does it require that they all be able to survive for several months after the outbreak of a major war without suffering any losses, deliberate or inadvertent, through antisubmarine warfare. The principle of deterrence would seem to be adequately served if ballistic-missile submarines on station could survive a full first-strike onslaught for a few hours in order to fire their weapons and if submarines not on station but in sanctuary or in the broad ocean reaches had a good probability of surviving a prudent trip of five or 10 days to firing position.

If war should break out between the U.S. and the U.S.S.R. and if nuclear weapons were used only at sea, the major consequence would be the destruction of each country's relatively few capital surface ships. Submarines, by and large, would survive. Although one cannot exclude the use of nuclear weapons against missile-carrying submarines, they do not change the conclusions. I shall argue the compatibility of an invulnerable submarine-based deterrent force with the active conduct of a tactical, nonnuclear and antisubmarine war.

Antisubmarine warfare involves at least four activities that proceed in logical sequence: intelligence, detection, localization and destruction. An optional tracking stage of hours or days can be imagined between localization and destruction. If the antisubmarine weapon is an antisubmarine mine, of course, no tracking is involved and the other activities are collapsed in time (although potentially intelligence can be fed into the mine from outside in advance or at arbitrary intervals).

Intelligence includes such information as the number of enemy submarines of each class, for example attack submarine (SS), nuclear-powered attack submarine (SSN), ballistic-missile-launching submarine (SSB), nuclear-powered ballistic-missile-launching submarine (SSBN) or cruise-missile-launching submarine (SSG or SSGN). Other kinds of information include number and range of missiles, number and type of torpedoes, speed and endurance, noise level, and sonar and radar capability. In addition to this intelligence on force levels and technical characteristics there is tactical intelligence, dealing broadly with the number of enemy submarines out of port, the number in each ocean basin, operating tactics, special vulnerabilities and so on. Tactical intelligence shades into detection and localization when a submarine can be limited to a certain area.

Over the years various methods have been employed for detecting submarines, including the detection of periscopes by radar, radio-frequency direc-



ADVANCED ANTISUBMARINE HELICOPTER, designated the SH-3H, is viewed from below in this schematic diagram. The helicopter is equipped with five different submarine-detection systems and is armed with a pair of lightweight 12-inch Mark 46 active acoustic homing torpedoes. The first eight SH-3H helicopters, built by the Sikorsky Aircraft Division of United Aircraft Corporation, are scheduled to be delivered to the U.S. Navy in mid-1972.

tion finding and visual sighting from aircraft. Although all play a role in tactical antisubmarine operations, the only currently effective way to locate a submerged submarine from any distance is by the use of acoustic methods, which can be either passive or active. Passive techniques include dispersed or concentrated arrays of fixed hydrophones, sonobuoy fields and hydrophones mounted either on surface vessels (usually destroyers) or on other submarines. Active techniques all require some kind of sonar gear: a sound generator combined with a hydrophone to receive reflected sound waves. Sonar systems can be mounted on submarines, surface ships or on devices towed at various depths from ships or helicopters. In addition sonar gear can be deployed in fixed underwater systems.

Passive techniques rely simply on the detection of noise generated by a submarine. When a submarine is running at high speed, the principal sound energy detectable at long range is radiated primarily by the propeller and by water flowing around the hull of the ship. At low speeds the major source of submarine noise is sound generated by installed machinery of various kinds, particularly rotating machinery that is unbalanced. Even a very noisy submarine, however, radiates less than one watt of acoustic power. One can therefore imagine the problem of detecting this tiny signal at long range in the presence of noise from waves and wind, undersea life and hundreds—even thousands—of surface vessels.

The propagation of sound in the ocean is extremely complex. Since the velocity of sound under water varies with temperature, pressure and salinity, it also varies with depth, position and season. These variations in sound velocity under some conditions bend underwater sound waves as much as 15 degrees, so that there is no guarantee that noise radiated from a submarine will consistently reach even the most sensitive passive listening post. On the other hand, sound waves are guided for thousands of miles with surprisingly little loss if they happen to be generated in the deep sound channel, a plane of minimum sound velocity lying at a depth of some 4,000 to 6,000 feet. A sound signal can travel 3,000 miles in the deep sound channel and still retain 1 percent of the intensity it had after traveling only 30 miles (suffering only cylindrical spreading loss). The existence of the deep sound channel is of no direct help in submarine detection, however, because existing submarines cannot safely descend to that depth and would be unwise to do so even if they could. When a near-surface hydrophone is moved away from a near-surface submarine sound source, the received signal often drops rapidly beyond a few kilometers and then rises to a high level again in successive "convergence zones" at multiples of about 50 kilometers (the distance required for a sound wave initially horizontal to be refracted through and beyond the deep sound channel and to return again to the surface).

Detecting a submarine in the immense background of ocean noise requires discrimination or processing gain. The desired source is known to be a point within a few hundred meters of the surface. Moreover, the source can be assumed to have a stationary spectrum in which there are usually strong line components. For decades large cadres of competent technicians have been working on the problems of extracting the submarine's "signal" from the noise in which it is immersed; the result has been a large variety of systems emphasizing different choices of available parameters, guided by operational considerations, analysis or tradition.

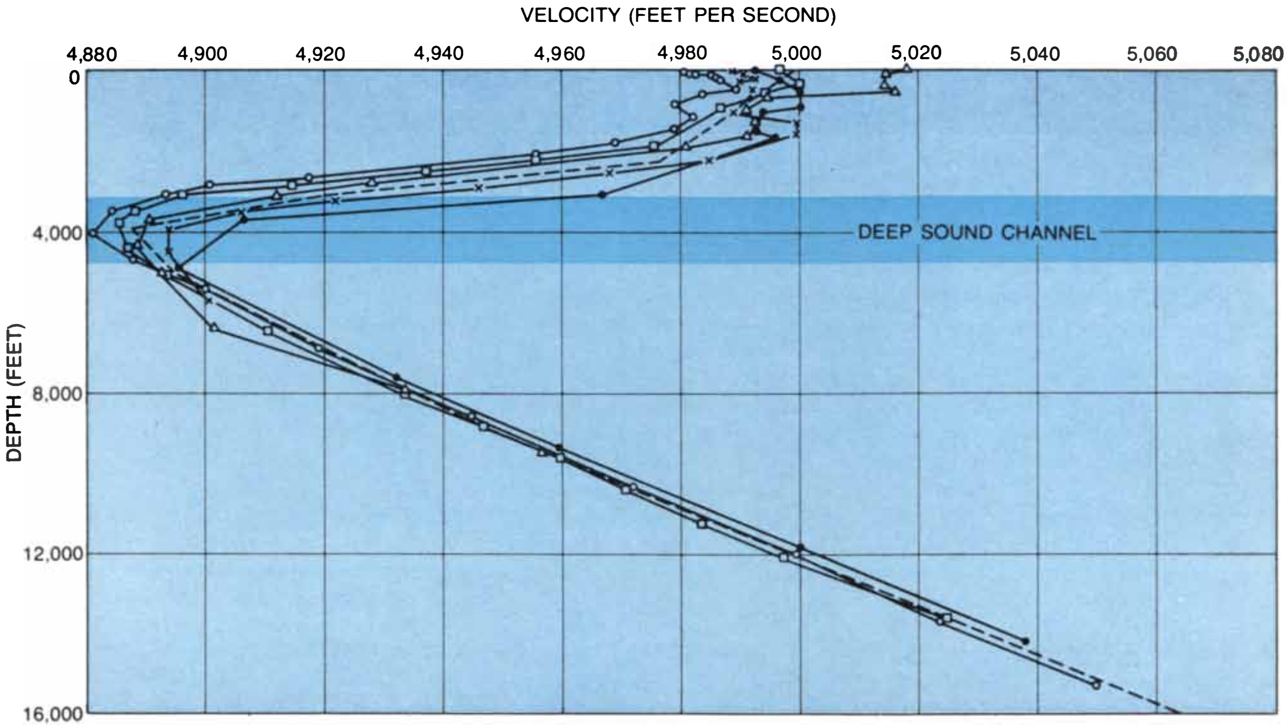
Basically the signal received from a source at a given assumed point can be enhanced with respect to noise or other point sources by adding in a coherent way the signals received at hydrophones located in different places. In a fixed dispersed array hundreds or thousands of hydrophones could be distributed throughout an entire ocean basin. The signals detected by the various instruments could be fed into a large computer provided with a program that assumes that a submarine, located somewhere in the detection area, successively occupies many possible points; for each point the program could introduce an appropriate time delay in the electrical output from each hydrophone. The computer could then add delayed electrical signals and ask whether the resulting integrated signal resembles the sound from a submarine or is simply just noise.

In a fixed concentrated array the hydrophones might be spread in a line no more than a few hundred meters long. Here the computer tests the result of integrating the various signals after inserting time delays allowing for the sound source to lie in some n number of distinguishable directions in a 360-degree arc. With processing equipment of sufficient size and power one can arrange for the beams from n directions to be formed simultaneously rather than sequentially, thereby ensuring that no signal is lost. Both dispersed and concentrated fixed arrays can incorporate hydrophones at various depths; by introducing vertical directivity the signal from distant sources is enhanced and locally generated surface noise is reduced.

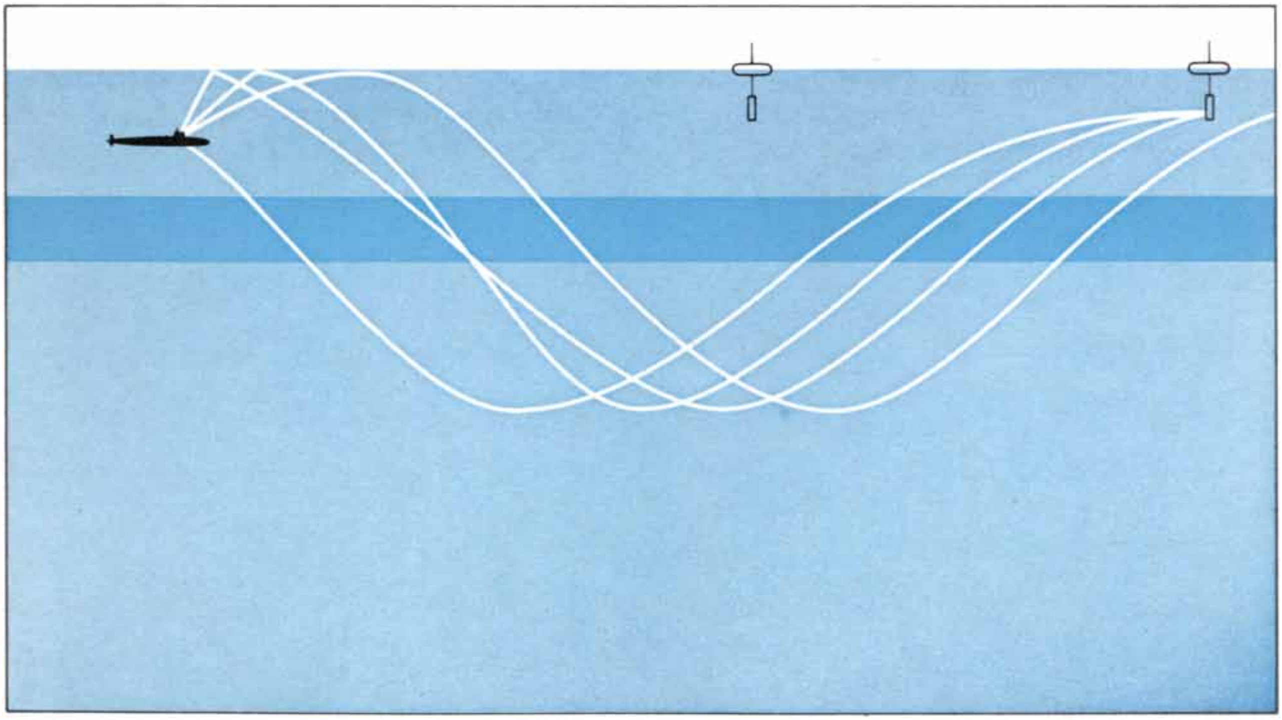
Sonobuoy fields are most expeditiously sown from and monitored by aircraft. The U.S. has both shore-based aircraft (the P-3C) and sea-based aircraft (the S-2 and soon the S-3) equipped to drop sonobuoys weighing a few kilograms from either low or high altitude. A hydrophone, dangling from each sonobuoy at a substantial depth, picks up pressure signals that are then relayed by radio to the aircraft. In principle the signals could be added coherently, as are the signals from fixed arrays, but this would be difficult with sonobuoy fields since the buoys have considerable freedom of movement. The alternative is to monitor the field by listening to one sonobuoy after another to determine whether or not a submarine is in the vicinity. Processing gain can still be obtained from a single hydrophone by making a time or frequency analysis of the sound signal. An aircraft will have a search rate of so many thousand square kilometers per hour, limited not so much by its flight speed as by its ability to monitor each sonobuoy effectively.

The most highly regarded vessel for passive acoustic detection of submarines is another submarine, since it is far quieter than a surface ship. Moreover, a hunter-killer submarine is large enough to carry an array of hydrophones to produce a narrow listening beam for long-range detection. Depending on the sound-velocity profile and the noise level of the hunted and the hunter, the effective range for passive detection can vary from less than a kilometer to 100 kilometers or more. Factors affecting range include the reflection of sound waves from the ocean surface and ocean bottom and the signal-processing capability of the hunter submarine. As might be expected, passive acoustic measures become less effective as submarine designers learn to build quieter vessels and submarine commanders learn how to minimize the radiation of noise. In any event a submarine lying "dead" in the water cannot be detected by passive techniques. As a result the U.S. and the U.S.S.R. both rely heavily on active acoustic-detection, or sonar, systems.

Although sonar gear is installed routinely on both surface vessels and submarines, one of the potentially most effective antisubmarine measures makes



VELOCITY OF SOUND IN THE OCEAN varies with temperature, pressure and salinity, and hence also with depth, as can be seen from this record of a series of measurements made at several locations in the North Atlantic Ocean. As a consequence of the complex nature of underwater acoustic conditions there is no guarantee that noise radiated from a submarine will consistently reach even the most sensitive acoustic-detection device. Sound waves generated in the deep sound channel, a plane of minimum sound velocity lying at a depth of approximately 4,000 feet in this part of the Atlantic (*dark colored band*), can be guided for thousands of miles with little loss, but this is of no direct help in submarine detection, since existing submarines cannot operate at that depth.



"CONVERGENCE ZONES" are observed when a near-surface acoustic detector is moved away from a near-surface submarine sound source. The received signal often drops rapidly beyond a few kilometers and then rises to a high level again at multiples of about 50 kilometers. The successive intervals that mark the convergence zones correspond to the distance required for a sound wave initially horizontal to be refracted through and beyond the deep sound channel and to return again to the surface of the ocean.

use of helicopters operating from carriers, destroyers or other sea platforms. The sonar devices are lowered from the helicopter by winch and can descend to depths of several hundred meters. An explosive charge can be detonated to produce an omnidirectional sound wave, or magnetostrictive or piezoelectric systems can be used to generate narrow beams of sound. The sound-wave probe is reflected from the sea surface, the sea bottom and marine animals, large and small, as well as from any submarine in the vicinity. The reflected signals are picked up and transmitted by way of the hoist cable to the helicopter for processing and display. To achieve efficient radiation and useful angular discrimination from helicopter-suspended platforms one uses sound frequencies in the range between one and 10 kilohertz (1,000 and 10,000 cycles per second) as compared with the subkilohertz waves employed for long-range detection. Signals in the range of 100 kilohertz to one megahertz attenuate so rapidly and become so scattered that they are not useful for submarine detection at long range.

In principle active sonars can detect moving objects more readily than stationary objects, because the Doppler shift in the sonar echo can be used to enhance the signal. On the other hand, even when the acoustic energy is confined to an ocean layer including the submarine, active sonars suffer a two-way loss: the power illuminating the target falls off directly with the distance, which means that the return signal decreases as the square of the distance. In other words, for equivalent received signals an active sonar must illuminate the target with at least as much energy as the target is radiating by itself (assuming that it is radiating and not just lying dead in the water), although detection can be accomplished with weaker signals because the precise form of the signal is known and often chosen to be best adapted to signal processing.

The purpose of active sonar is not necessarily to sweep the oceans clear of submarines; such systems are often employed to provide a submarine-free region a few kilometers in diameter in which merchant or military ships can be free from attack. When submarines had only unguided torpedoes to attack shipping, active sonars mounted on destroyers or suspended from helicopters provided reasonable protection. Now, however, submarines carry acoustic homing torpedoes with a range of 10 kilometers or more. And the cruise missiles carried by many Russian submarines have an effective range of hundreds of kilometers, which means that active sonar cannot provide a point defense against such submarines.

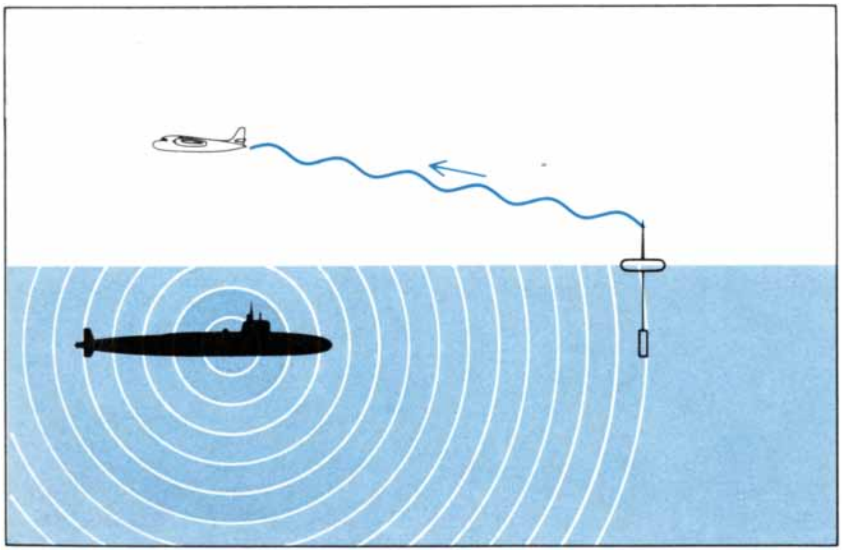
In the past decade much work has been done on active sonar systems anchored to the ocean bottom in an effort to provide detection of submarines at ranges of hundreds of kilometers. These fixed-active systems, as they are called, emit a narrowly focused beam of sound at frequencies of a few hundred hertz and a power level of several megawatts. Such installations not only are expensive but also have the drawback of broadcasting their location, which makes them vulnerable to countermeasures. The opposition need not even destroy the installation; by sampling the pulses emitted by the system it can contrive to "spoof" or jam it.

An alternative to the fixed-active sonar system is the semiactive, or bistatic, system, in which the transmitters and receivers are widely separated. Such systems are probably harder to spoof or jam than ordinary active sonars and they provide more leeway in the depth at which the transmitter can be located. At very high power levels a deep transmitter is less limited by cavitation: the rapid expansion and collapse of bubbles that arise from gases dissolved in the water.

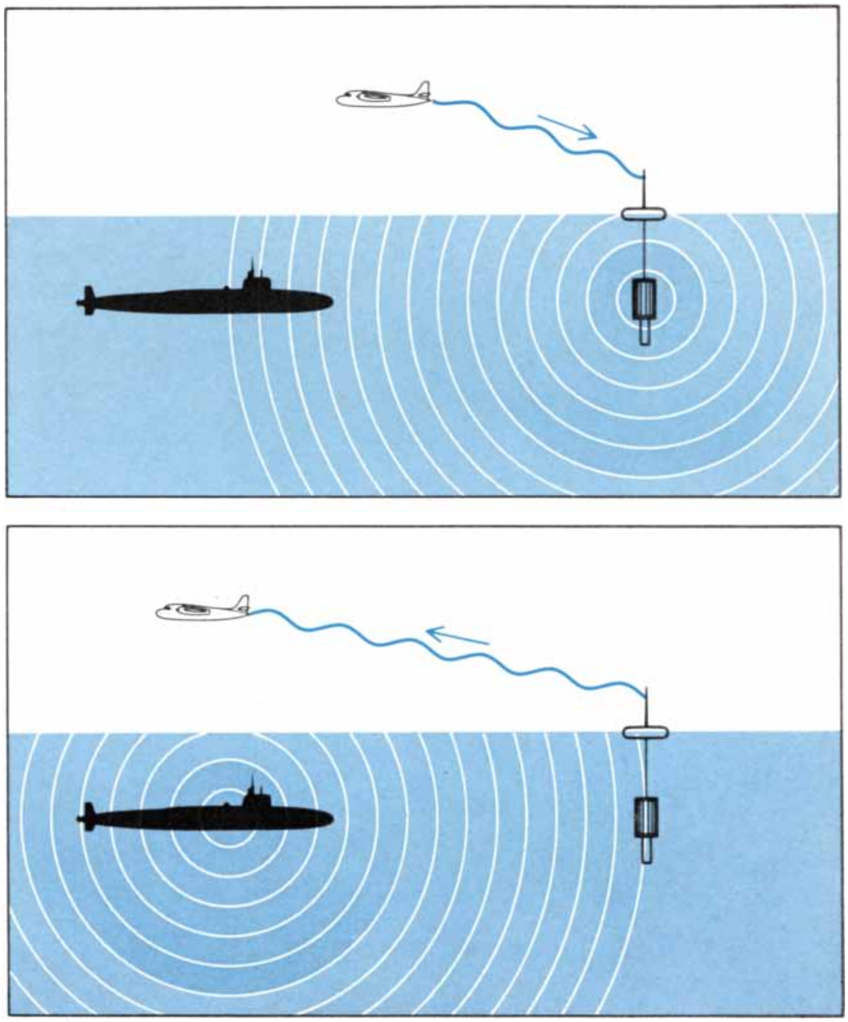
After a submarine has been detected in a certain target area, say an area 50 kilometers square, the next step is to pinpoint its position. Localization can be done from the air, from the surface or from below the surface. If an aircraft such as a P-3 or an S-2 is employed, it can drop directional sonobuoys that are capable of indicating the bearing from the sonobuoy to the submarine. The aircraft then follows along this line by dropping more sonobuoys. When the submarine has been located to an accuracy of one kilometer or so, its position can be determined still more precisely by low-level passes with a magnetic-anomaly detector (MAD) to sense the presence of a large mass of submerged steel.

If localization is conducted by helicopter, the preferred detector is active sonar suspended from the aircraft, which can determine range accurately and bearing somewhat less accurately. If one assumes that the sonar has a range of three kilometers and that five minutes of listening are needed each time the sonar unit is lowered into the water, one can estimate that one helicopter can search 300 square kilometers per hour. Thus within an hour six helicopters should have a 50 percent chance of finding a single submarine known initially to be in a square area 50 kilometers on a side.

If the search is conducted by surface ships, active sonar is used in essentially the same way as by helicopters except that a surface ship does not have nearly



PASSIVE ACOUSTIC DETECTION of a submerged submarine relies on picking up the noise radiated by the submarine itself. In this case the passive acoustic detector is a hydrophone dangling from a buoy at a substantial depth; signals detected by the hydrophone are relayed by radio to the aircraft for analysis. Hydrophones of this type and size can be sown by aircraft in either dispersed or concentrated arrays. Larger hydrophones can also be mounted on the hulls of surface vessels (usually destroyers) or on other submarines.



ACTIVE ACOUSTIC DETECTION of a submerged submarine requires a sound generator combined with a hydrophone to receive reflected sound waves. "Sonar" systems of this type suffer a two-way loss: in a near-surface sound channel the power illuminating the target (*top*) falls off directly with the distance, which means that the return signal (*bottom*) decreases as the square of the distance. Deep or bottom-mounted sonars suffer spherical spreading of the sound energy; hence the illuminating power falls off as the square of the distance, and the total energy received at the sonar falls off as the fourth power. Besides being deployed in aircraft-sown sonobuoy fields, sonar systems can be mounted on other submarines, surface ships or on devices towed at various depths from ships or helicopters.

the speed advantage over a submarine that a helicopter has. This disadvantage is partly compensated for by the surface ship's ability to carry a sonar of much higher power and effectiveness than the equipment a helicopter can carry.

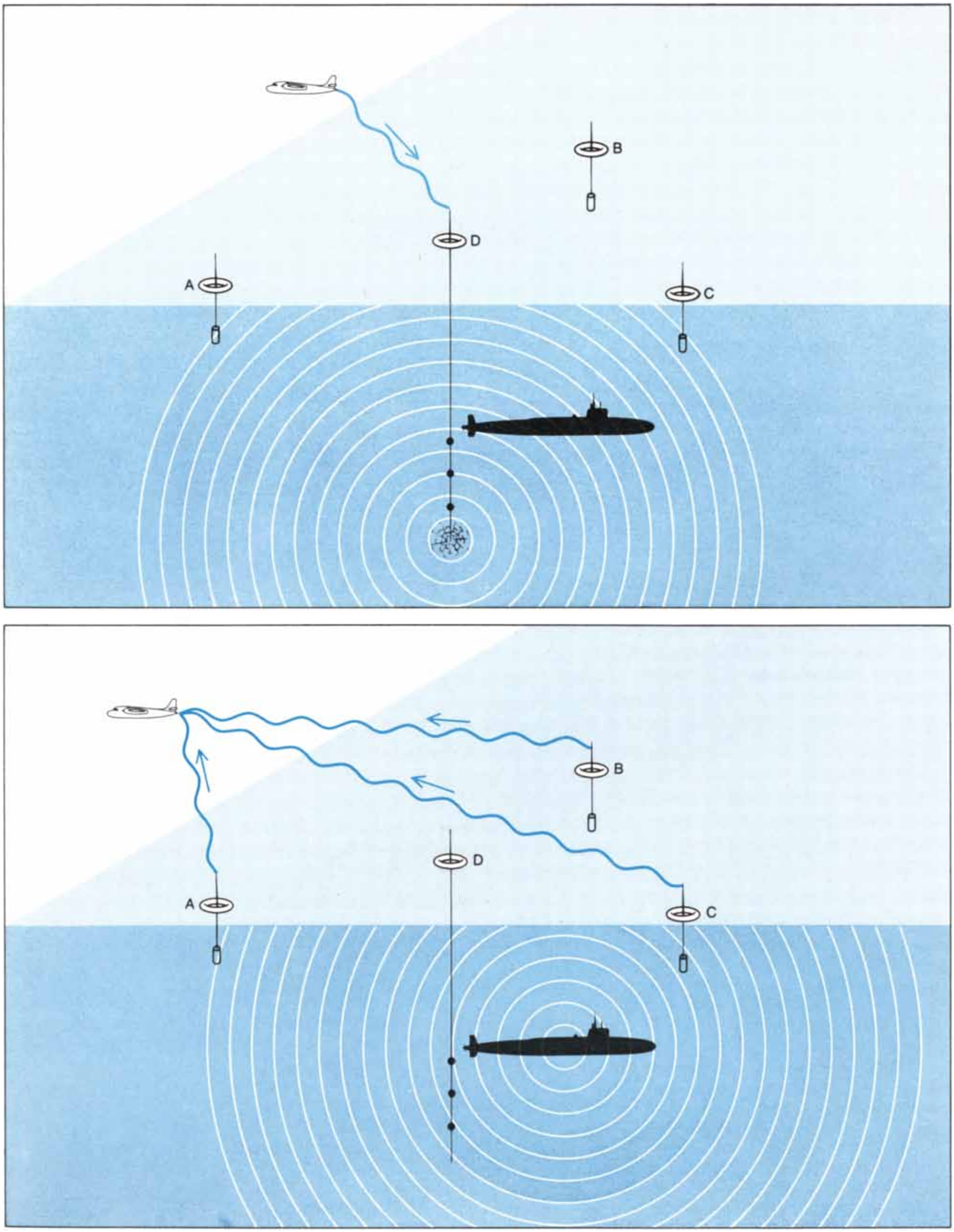
Underwater search by one or more hunter-killer submarines is perhaps the slowest but not necessarily the least effective way of localizing a submarine. A hunter-killer submarine has an optimum search speed of between five and 15 knots. For a given sonar range the area searched increases linearly with speed, but the self-noise generated as speed increases reduces the effective sonar range. Hence the optimum search speed is a compromise.

In antisubmarine-warfare exercises an "enemy" submarine that with much effort has been detected and localized is often tracked for hours or even days. The tracking skill thus developed in peacetime, however, has little applicability in wartime. Under wartime conditions an enemy submarine would ordinarily be sunk as soon as it was localized. As for tracking the other side's ballistic-missile submarines 24 hours a day in peacetime, the simultaneous tracking of 30 to 50 modern submarines is so difficult that its feasibility seems doubtful. Even if the job could be done, it is the thesis of this article that continuous tracking is distinctly undesirable if one concedes that the nuclear deterrent is valuable and should be preserved.

Let us see nonetheless what continuous tracking involves. Antisubmarine warfare is the primary mission of the U.S. fleet of nuclear-powered hunter-killer submarines, numbering some 57 vessels. Their chief sensor is low-frequency passive sonar. Depending on the state of the sea, they are probably able to detect a moving Russian submarine at distances ranging from a few kilometers to perhaps 20 or 30 kilometers under more favorable conditions. An effort to carry out passive acoustic tracking reliably at a range of 10 kilometers or more in the face of changing sea conditions, normal variations in sound velocity, sea noise and the evasive measures readily available to the quarry seems doomed, however, to failure. If such tracking could be accomplished at all, several hunter-killer submarines would probably have to be assigned to each hunted submarine. Moreover, if the necessity should arise, the destruction of an enemy submarine at a range of 10 kilometers or more before it could fire its missiles might call for a weapon faster than the standard torpedo. For these and other reasons, including the possibility of acoustic countermeasures, my own view is that effective covert tracking is not a plausible disarming threat to the submarine-based deterrent forces of either the U.S. or the U.S.S.R.

Active sonar tracking from submarines, surface vessels or aircraft is another story. Given the mission to maintain at all times the ability to destroy the enemy fleet of ballistic-missile submarines before it could launch its weapons, one would immediately think of active trailing from extremely short range, say a few hundred meters. At that distance a sonar operating at a frequency of 100 to 1,000 kilohertz would be compact and low-powered and could provide a clear pictorial representation of the target. For example, a 1,000 kilohertz (one megahertz) sonar would provide a resolution of 30 centimeters on a target 200 meters away. Mounted on a specialized trailing vehicle (submerged, on the surface or airborne), such a sonar could provide a detailed picture of its quarry every few seconds. The tracker could therefore sail in formation with the quarry without fear of collision.

This is not the place to describe the command-and-control arrangements that would enable the small specialized tracking ship to destroy its quarry on



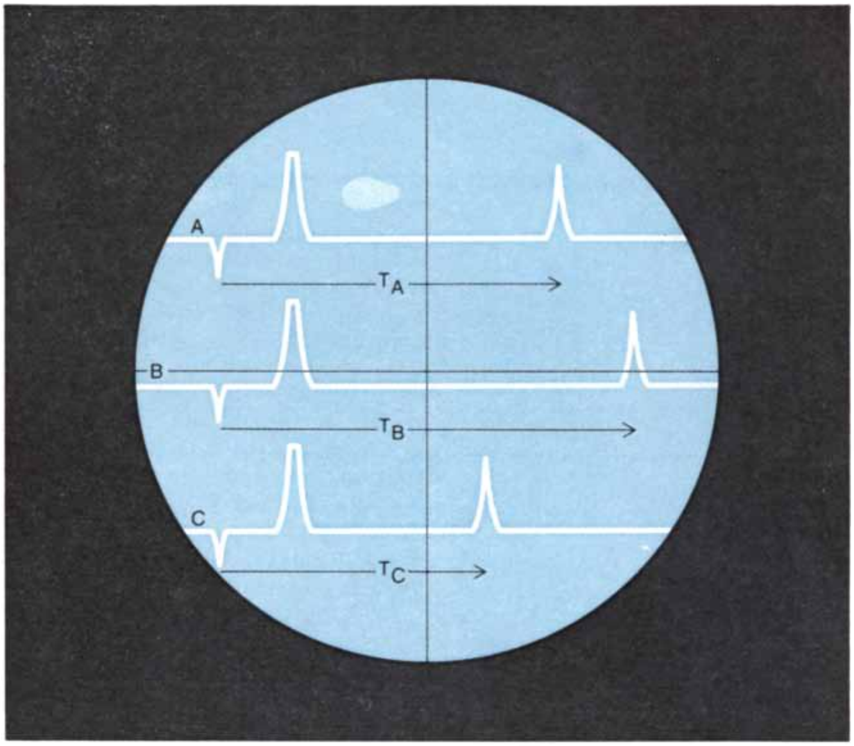
SEMIACTIVE ACOUSTIC DETECTION involves systems in which the transmitters and the receivers are widely separated. In the idealized system portrayed in the pair of diagrams on this page an ASW aircraft has sown a field of passive hydrophones in the vicinity of a submerged submarine. A string of explosive charges is suspended from a similar aircraft-delivered buoy to a depth of several hundred meters. The charges are detonated singly on command to produce a series of omnidirectional sound waves (*top*). The reflected signals, or echoes, are picked up by the hydrophones and transmitted to the aircraft for processing and display (*bottom*).

command or that would lead to the destruction of both ships if the quarry should attempt to use force against the tracker. The important fact is that the characteristics of such a close-in tracking ship (be it a submarine, a surface vessel or an aircraft) are quite different from those of any craft deployed in large numbers today. The craft would need to have endurance and good seakeeping qualities; it would have no need for more than a few torpedoes or other destructive weapons, and none for a large crew or great flexibility of operation.

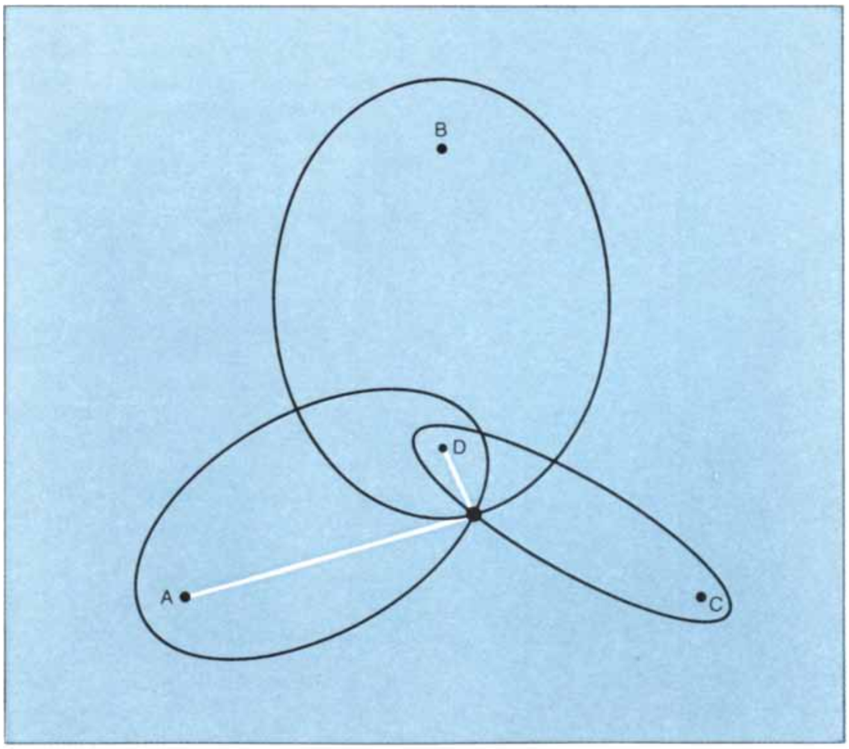
The chief difficulty in this type of active tracking seems to lie not in maintaining tracking but in ensuring that a tracker is assigned to every ballistic-missile-launching submarine that leaves the home port and crosses a detection barrier of some kind. Obviously the other side will seek to avoid detection by employing various stratagems, perhaps by sending its missile submarines across the barrier in clumps or by mixing them with non-missile-submarine decoys. The many possibilities need not concern us here. As I have argued above, tracking (particularly active tracking) has little application to tactical antisubmarine warfare, but if done successfully and without opposition, it could jeopardize the survivability of the ballistic-missile-submarine forces of the U.S. and the U.S.S.R.

The U.S. Navy has two principal weapons for destroying enemy submarines: the lightweight 12-inch Mark 46 active acoustic homing torpedo and the full-size 21-inch Mark 48 active-passive acoustic homing torpedo (which is just coming into service). The 12-inch torpedo is essentially the only effective antisubmarine weapon available to aircraft and helicopters. It can also be launched from the deck of a ship or delivered by a drone helicopter. A rocket-boosted version has more range and can get to its target faster than the standard Mark 46. The Mark 48 is a long-range torpedo with a large warhead; when it is launched by a hunter-killer submarine, it can be guided electrically by means of a control wire. Like any other weapon, a homing torpedo will not be perfectly reliable in actual combat, and it may be vulnerable to countermeasures.

Specialists in antisubmarine warfare have long been attracted to more effective antisubmarine mines. Such mines could be planted defensively to protect harbors or to cordon off a landing area for an expeditionary force. They could also be used to keep enemy submarines



RECEIVED SIGNALS from the semiactive detection system shown in the illustration on the opposite page are processed on board the monitoring aircraft and displayed on a cathode ray tube. The times at which the signals are received from the various hydrophones are then used by an operator to plot the location of the submarine (*see illustration at bottom*).



SUBMARINE IS LOCATED on the basis of the times at which the hydrophones picked up the signals reflected from the submarine. The observed time interval T_4 between the explosion at D and the reception of the echo at sonobuoy A defines an ellipse with foci at A and D. (The sum of the distances from any point on an ellipse to the two foci is constant.) The information from each hydrophone is used to draw an ellipse on which the submarine is situated. The intersection of the three ellipses indicates the exact location of the quarry.

from leaving their home ports, to create a barrier across straits or even to limit travel in the open ocean.

Since ordinary mines have a limited radius of effectiveness (submarines could pass above or below them), such mines can be used against submarines only in shallow water. There have been recurrent proposals, however, for a mine that would either drive itself downward when a trip wire indicated a submarine was passing below it or drive itself upward when a submarine was detected above it. Mines with this capability are probably under development in both the U.S. and the U.S.S.R.

If such a mine were based on the Mark 46 torpedo, it should cost well under $100,000 and have a 50 percent chance of destroying any submarine passing within one kilometer of it. Thus 500 mines would be enough to cover the gap of 1,000 kilometers between Greenland and the British Isles. Actually the mines would not be sown cheek by jowl in a thin line; they could better be deployed in a minefield 100 kilometers deep with a density of one mine per two kilometers of front so that 500 mines would still suffice.

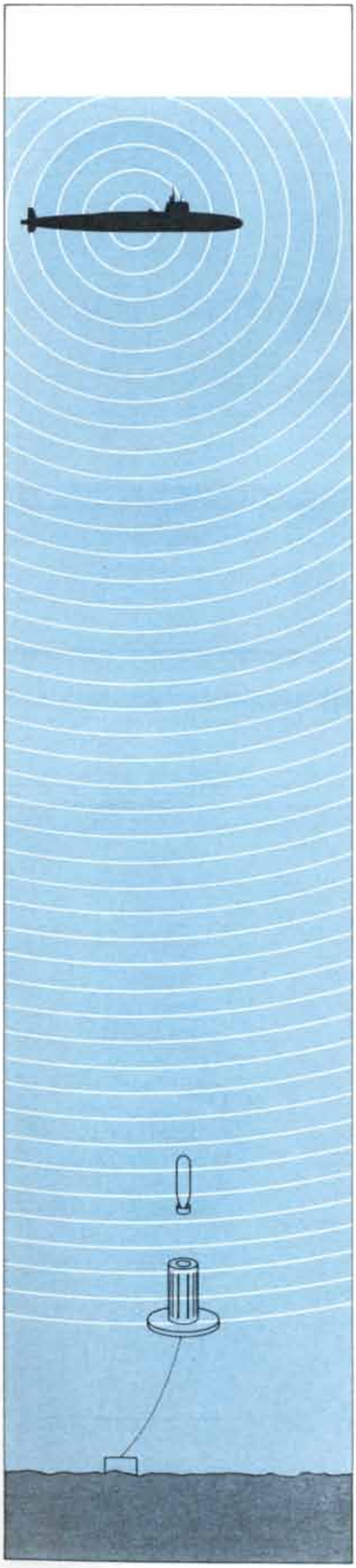
The mines would allow surface ships to pass without harm, and the commanders of friendly submarines could traverse the minefield safely with the aid of charts. The mines could be remotely activated and deactivated. They could be laid by aircraft, surface ships or submarines. For about $50 million (less than the cost of one destroyer) it should be possible to provide a robust, effective barrier 1,000 kilometers long with low physical and political vulnerability.

Let us now try to distinguish between the kind of antisubmarine force needed to protect surface shipping and the kind of specialized capability needed to find, track and destroy on signal ballistic-missile submarines. Because these two needs impose conflicting requirements on the design of antisubmarine-warfare forces it is not surprising that the U.S. Navy sometimes takes a self-contradictory position when it tries to argue on the one hand the invulnerability of its Polaris-Poseidon submarine fleet and on the other the effectiveness of its antisubmarine forces. I shall argue that one can develop antisubmarine measures sufficient to protect surface shipping during an actual conflict without developing an antisubmarine capability of such a nature that it jeopardizes the other side's submarine-based nuclear deterrent.

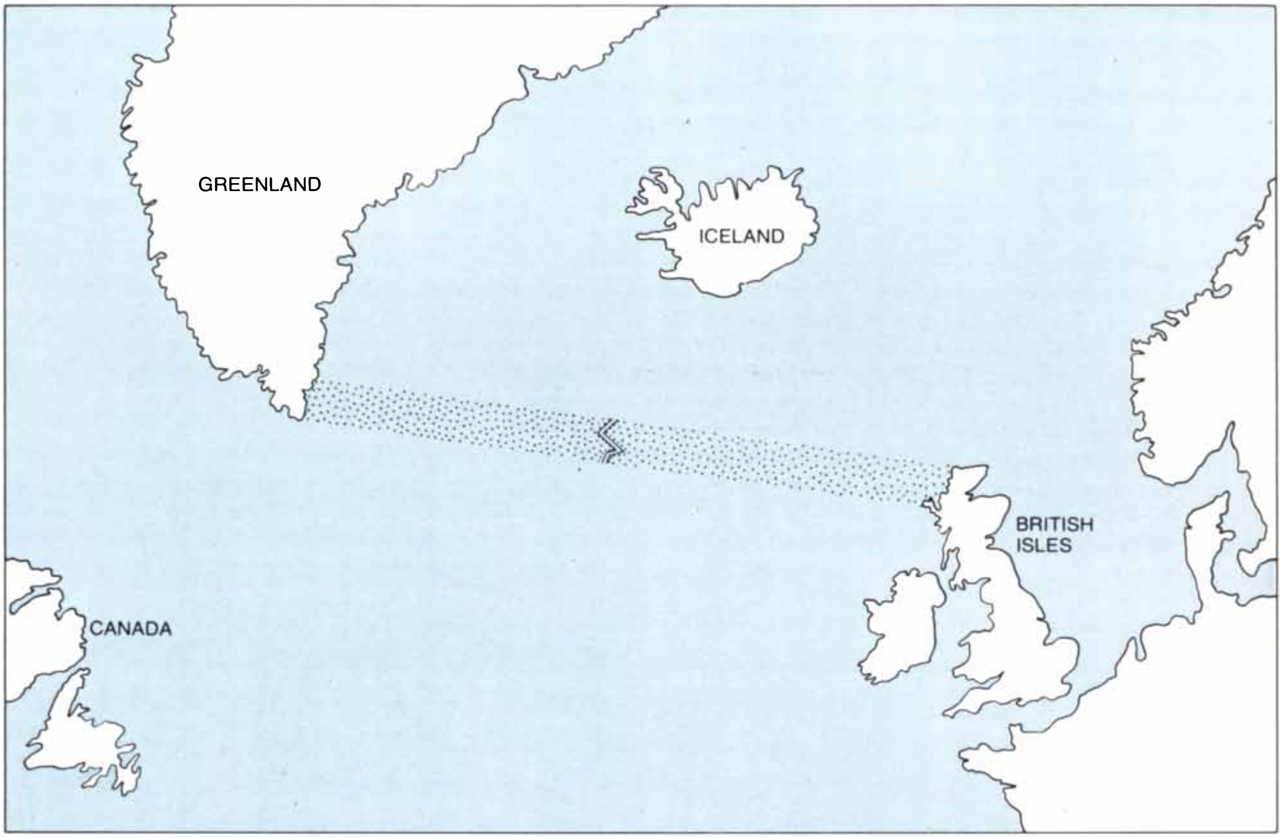
Unlike major cities or even aircraft carriers, merchant ships are plentiful and of modest value; they are substantially less costly and easier to replace than nuclear-powered attack submarines. A merchant ship delivered in 1972 with a speed of 22 knots and a deadweight of 29,000 tons costs about $23 million. A U.S. nuclear submarine costs more than $150 million. The average crew size of nuclear submarines also substantially exceeds that of modern merchant ships. The U.S. has some 100 attack submarines; the U.S.S.R. has about 300 oceangoing submarines of all types. During World War II Germany lost about 700 submarines and sank 2,600 merchant ships, accounting for 13 million tons of shipping in the North Atlantic. Therefore in round numbers Germany lost one submarine for every 17,000 tons of Allied shipping sent to the bottom. If in another war U.S. antisubmarine forces were able to destroy one enemy nuclear-powered attack submarine for every 10 merchant ships sunk, we should probably regard the exchange ratio as being in our favor. In such a war of attrition the nonnuclear-powered submarine could be the greater threat.

A well-protected convoy remains a principal means for protecting surface vessels. With the various antisubmarine measures described above it should be possible to achieve a kill ratio of one enemy submarine for every 10 merchantmen sunk. Although this ratio might be acceptable for ordinary military and civilian cargoes, it would not be satisfactory if the surface vessels carried troops or critical materiel required at a specific time by troops being ferried by air.

Protection for such vessels is ordinarily imagined to be provided by destroyer-escort vehicles with extra-powerful sonars that provide a moving screen extending well beyond torpedo range and within which enemy submarines would have a high probability of being located and destroyed. In my own opinion the best way to create such a screen would be to improve the capability of antisub-



ANTISUBMARINE MINE of a type that is probably under development in both the U.S. and the U.S.S.R. would have a much greater radius of effectiveness than ordinary mines. The hypothetical device shown incorporates an acoustic homing torpedo, which is designed to drive itself upward when a submarine is detected above it. Such a mine would cost well under $100,000 and have a 50 percent chance of destroying any submarine passing within one kilometer of it.



ANTISUBMARINE MINEFIELD based on the use of mines such as the one shown on the opposite page could be laid to create a barrier across the 1,000-kilometer strait between Greenland and the British Isles. For a minefield 100 kilometers deep with a density of one mine per two kilometers of front 500 mines would be needed. The mines would allow surface ships to pass without harm, and friendly submarines could traverse the minefield safely along mine-free lanes with the aid of charts. The mines could be remotely activated and deactivated. The production cost of such an ASW measure (about $50 million) would be less than the cost of one destroyer.

marine helicopters by placing the analysis and data-processing equipment for each helicopter in a van on the deck of a ship rather than in the helicopter itself. With the room saved each helicopter could be provided with several active sonars, which could be deployed by a single helicopter in leapfrog fashion, with the control and signal analysis transmitted by radio between ship and sonars. In this way it should be possible to increase the effectiveness of antisubmarine helicopters by a factor of five, while reducing the crew from four to no more than two persons. Although such improved helicopters can provide a potent local defense, a ballistic-missile-launching submarine could easily move out of range when it detected the approach of such a convoy.

Because a convoy has certain inherent vulnerabilities, such as inviting attack by straight-running torpedoes or by cruise missiles launched from a considerable distance, it is sometimes preferable to allow fast merchant ships to sail independently at maximum speed. The threat to such ships from torpedo attack might be reduced to negligible proportions if each ship carried some retaliatory homing torpedoes that would be launched automatically from the deck of the ship if the ship were struck by a torpedo. This tactic could deter torpedo attack (without defending against it) in just the same way that the strategic offensive force reliably deters a first strike.

The defense of aircraft carriers presents a special problem. An aircraft carrier can scarcely be hidden. Its screws can be heard all the way across an ocean; it is highly visible on radar and can be detected by other means. For these reasons carriers are built to resist damage; as many as a dozen torpedoes are said to be needed to sink one. A close-in defense against torpedo-launching submarines can be provided by the same helicopter-sonar screen already described for the protection of convoys. As some additional protection against cruise missiles launched from a distance of hundreds of kilometers carriers are already equipped with rapid-firing automatic weapons that provide a last-ditch defense. These defenses against cruise missiles could be further improved. One should probably consider installing active antitorpedo defenses directly on the carrier's hull. None of these measures for the defense of aircraft carriers poses any threat to the ballistic-missile submarine.

In any readily foreseeable war between the U.S. and the U.S.S.R. a major part of the conflict at sea will be asymmetrical: an essential mission of the U.S. Navy will be to keep the sea lanes open and an essential mission of the Russian navy will be to close them. Only in preserving the security of their submarine-based ballistic-missile forces are the missions of the two navies symmetrical. By emphasizing the antisubmarine measures discussed here, the U.S. can carry out its two missions; it should be able to protect surface shipping and nullify the Russian attack-submarine force without creating a credible threat to Russia's submarine-based nuclear deterrent. It is this freedom to pursue their individual and opposing interests, while not jeopardizing their common interest in an invulnerable sea-based deterrent, that could lead to arms-control agreements to provide enhanced and equal security for the SLBM deterrent forces.

As we have seen, neither country as yet has the means for negating that deterrent. Even if the U.S. were to improve its tactical antisubmarine techniques to the level I have described, they could not be regarded as a threat to Russia's ballistic-missile submarines. For example, antisubmarine measures that would require, say, 10 days to destroy half of the enemy's ballistic-missile submarines in the open ocean cannot be said to threaten the deterrent. If such measures required only 10 minutes, they would be a real threat.

Moving cordons of antisubmarine vehicles around convoys or aircraft carriers, designed to protect at most 1,000 square kilometers of ocean at a time, can be easily avoided by ballistic-missile-carrying submarines and hence present no threat to them. By the same token the statistical defense of individual merchant ships, triggered only by a torpedo explosion, is no threat.

Would the emplacement of advanced mine barriers be regarded as a threat? A barrier 1,000 kilometers long extending from Greenland to the British Isles could undoubtedly take a substantial toll of deterrent-force submarines that tried to cross it after hostilities began. We are assuming here that the conflict is originally nonnuclear and that both sides want to preserve their submarine-based deterrent to keep it that way. Even if the U.S. did not deliberately cooperate in helping deterrent-force submarines through the minefield, at least half should survive the one-way trip needed to reach firing position. The SLBM force could be stationed behind the minefield without suffering attrition; its deterrent power would not be impaired by the losses suffered on a one-way trip through a barrier if the order were given to move to firing position and launch. On the other hand, the minefield would take a prohibitive toll of antishipping submarines, which must make many trips in and out of their home port.

The question sometimes raised is: How should the U.S. respond if the Polaris-Poseidon fleet should suffer attrition during peacetime under circumstances that seemed suspicious? My personal view is that neither the U.S. nor

ASW AIRCRAFT CARRIER, the U.S.S. *Intrepid,* is capable of launching both ASW aircraft and ASW helicopters. The U.S. has four such carriers, the U.S.S.R. none. The photograph was made during an inspection cruise in the Atlantic Ocean in February, 1971.

the U.S.S.R. has any incentive to covertly reduce the size of the other's deterrent fleet. Furthermore, it is difficult to guarantee that such a covert action is not exposed. Finally, each side has equal vulnerability in its submarine-deterrent fleet.

A more difficult problem is to know how to conduct active, conventional ocean war, including tactical antisubmarine warfare, without an accompanying and unintended attrition of the opponent's submarine-based nuclear deterrent. Let us turn to that problem.

NEWEST ASW AIRCRAFT, the S-3A, was photographed during a test flight earlier this year. The jet-powered aircraft is scheduled to join the U.S. ASW forces in 1974, replacing the propeller-driven S-2. The carrier-based S-3A will carry a digital computer to enable its four-man crew to better analyze underwater sounds and other data. It will be armed with a variety of weapons: acoustic homing torpedoes, mines, depth charges, rockets and missiles.

The first step in reducing unintentional attrition would be formal recognition that it is in the mutual interest of the two countries to preserve the submarine-based deterrent not only in peacetime but also indefinitely during a conflict to forestall the escalation to all-out nuclear war. One can imagine many possible arms-control agreements that would then contribute to the continued invulnerability of the submarine-based deterrent.

Such an agreement might have two goals. First, it might prohibit either side from building or deploying a force that is technically capable of destroying the other side's submarine deterrent in a matter of minutes, hours or even a few days. The second goal, which might be more difficult to achieve, would be a set of measures that would reduce or perhaps eliminate the attrition of missile-launching submarines during a conventional war without restricting vigorous warfare to protect (or attack) surface shipping.

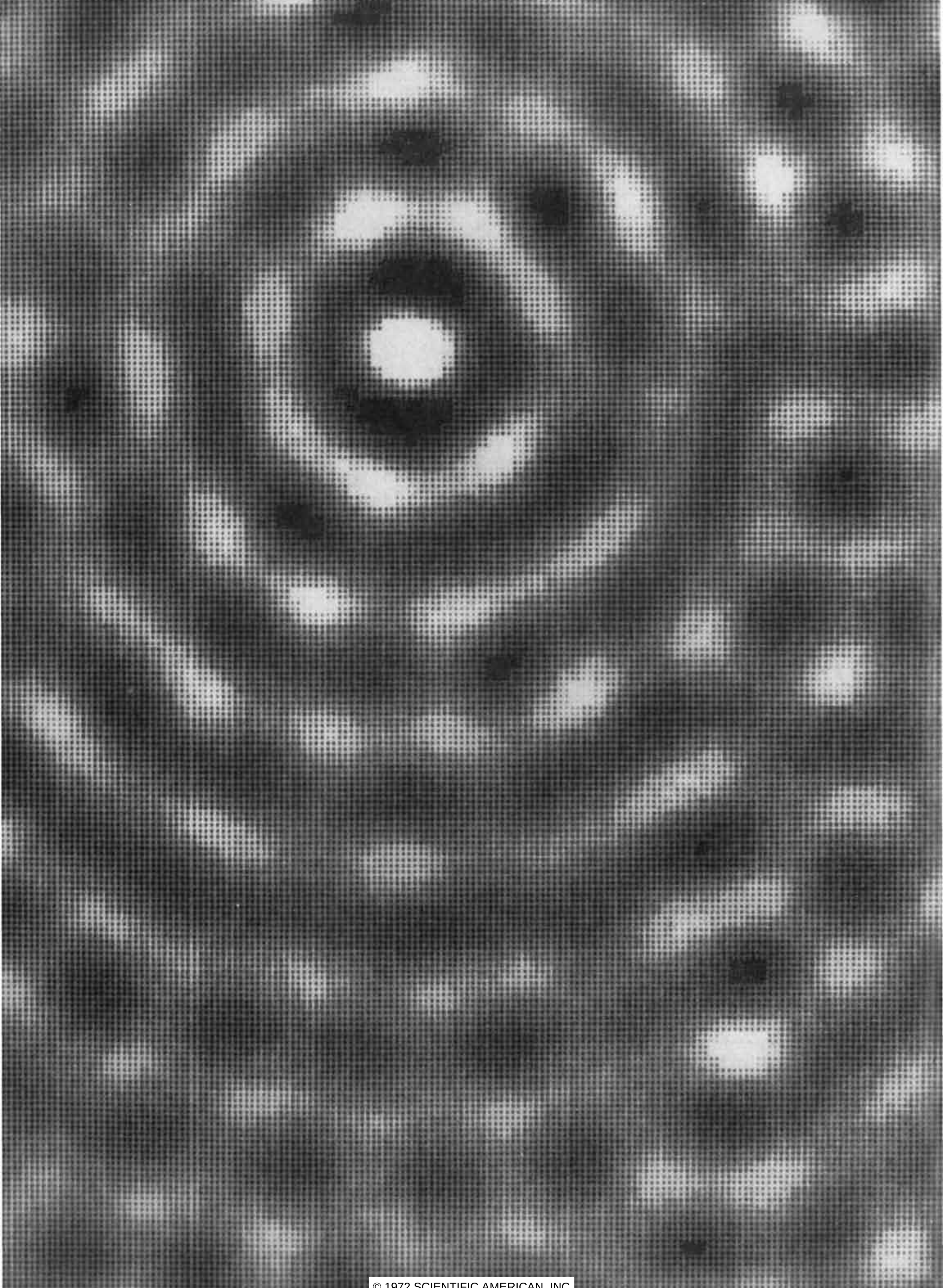
The first goal—the peacetime one—might be aided by prohibitions against the active tracking of missile submarines and by the creation of sanctuary areas. There are so many possible technical ways to discourage active tracking unilaterally that it is by no means assured that one side, invest what it will, could follow nearly all its opponent's submarines. It would nonetheless be valuable to reach an early agreement to eliminate fear of this threat, which seems to have no utility other than to threaten the survival of the strategic deterrent. Moreover, such an agreement seems readily verifiable. It should not be difficult for a submarine that is being tracked to recognize that it has (or has not) at all times a companion at a distance of a few hundred to a few thousand meters.

Additional protection for the submarine deterrent force in peacetime could be achieved by creating wide-ocean sanctuaries in which ballistic-missile submarines of one side or the other would be free to patrol, immune from surveillance of any kind. Clearly the sanctuaries must be large compared with the range of antisubmarine detection and attack devices, yet they must not be so large that they interfere with merchant shipping or with the transit routes of naval vessels, including deterrent-force submarines. Problems may arise in connection with third parties but these do not appear insoluble.

To limit the attrition of deterrent-force submarines during a conventional war, the U.S. and the U.S.S.R. might agree to provide safe-conduct routes through minefields or other barriers for each other's ballistic-missile submarines. The problem here is to distinguish between strategic missile submarines entitled to safe passage and attack submarines, which would still be fair game. The simplest solution would be to provide safe passage to any submarine that was effectively running on the surface and could thus be inspected.

Because surfaced submarines are not very good at traversing heavy seas, a submarine could be allowed to travel submerged provided that it towed a buoy emitting a distinctive radio signal. The signal could be monitored by surface ships, aircraft or even by satellites, and the position of such submarines could thus be known precisely and instantaneously. Buoy-towing submarines would be allowed to pass without harassment (although subject to surfacing on call) through clearly marked lanes in barriers; all other submarines would be subject to attack. The number of safe-conduct trips per month could be only enough to permit normal rotation of submarines in the deterrent force.

These three schemes only suggest the scope of possibilities for a SALT II agreement that would enhance the survivability of submarines designed specially for carrying long-range ballistic missiles. I should emphasize that in my opinion such submarines are probably adequately survivable in the absence of a new agreement. An important function of the agreement, however, would be to allay exaggerated fears that our Polaris-Poseidon fleet might suddenly be neutralized by a dramatic advance in the effectiveness of antisubmarine warfare. As a nonnegligible by-product of such an agreement both the U.S. and the U.S.S.R. would be enabled to reallocate either within or outside the military sphere much of the money that is likely to be spent in attempting to endanger the deterrent, if no agreement is forthcoming.



THE X-RAY SKY

More than 120 celestial X-ray sources are now known; the list includes at least one neutron star, one quasar, two galaxies, one double source and perhaps even a black hole

by Herbert W. Schnopper and John P. Delvaille

It has been 10 years since the discovery of the first celestial object (apart from the sun) known to emit X rays. A total of more than 120 X-ray sources have now been located and studied in some detail. The data come from nearly 100 rocket and balloon flights and from one major satellite experiment that has been in orbit for more than a year. The diverse nature of the experimental results has been a source of constant wonderment to the practitioners of X-ray astronomy. X rays have been observed from objects inside and outside the galaxy. Some of these objects are compact; others extend over large regions of the sky. At least one neutron star (the pulsar at the core of the Crab Nebula), one quasar (the object designated 3C 273) and two galaxies (the large and small clouds of Magellan) are known to emit X rays. The majority of X-ray sources, however, have not been identified with objects that emit either light or radio waves. Many X-ray objects vary in spectral type and total radiated power over a period of time. One source in Centaurus (Cen X-3) is believed to be an eclipsing binary: two objects revolving around a common center of gravity that eclipse each other as seen from the earth. This celestial zoo is an astronomer's delight, and each new experimental approach is quickly rewarded with a clearer look at the X-ray sky.

Extrasolar X-ray astronomy was born on June 18, 1962, when X rays from the direction of the constellation Scorpius were discovered from a rocket payload launched from the White Sands Missile Range. The experiment was designed by a group of workers at American Science and Engineering Inc. and the Massachusetts Institute of Technology. Although the location was imprecise, the experimenters were certain that an X-ray source existed in Scorpius, and they named it Sco X-1. Refinements were made to increase the sensitivity of the survey experiments that followed, and many new sources, among them the Crab Nebula, were discovered. The early positional information was too crude for a telescopic search to determine if there was a visual object that could be identified with each X-ray source. That situation changed quickly. A group of astronomers at the Naval Research Laboratory in Washington observed the Crab Nebula as the moon eclipsed it. From this occultation experiment they established that the X-ray-emitting region was about one minute of arc across. The American Science and Engineering–M.I.T. group measured the angular extent and determined the precise location of Sco X-1 during a rocket flight on March 8, 1966. These results immediately led to the discovery of an optical counterpart to Sco X-1: a faint blue star of the 13th magnitude.

The introduction of a special instrument called a rotating modulation collimator [*see illustration on page 30*] set the stage for the most precise location of X-ray sources. In a rocket experiment launched on October 2, 1969, five sources were located lying roughly in the plane of the galaxy near the galactic center. Although a careful search of the positions gave no indication of suitable candidates for optical objects, radio astronomers working with the facilities at the National Radio Astronomy Observatory in Green Bank, W.Va., found radio sources near these positions.

In addition to fairly compact discrete objects the X-ray sky is bathed with a background of low-energy X rays. The background spectrum was first noticed at higher energies. Its intensity rises sharply at lower energies, and it has been explored by several rocket experiments equipped with special detectors for low-energy X rays. The X-ray group at the University of Wisconsin attributes much of the background to activity within our own galaxy and has found a considerable contribution from a region called the North Polar Spur.

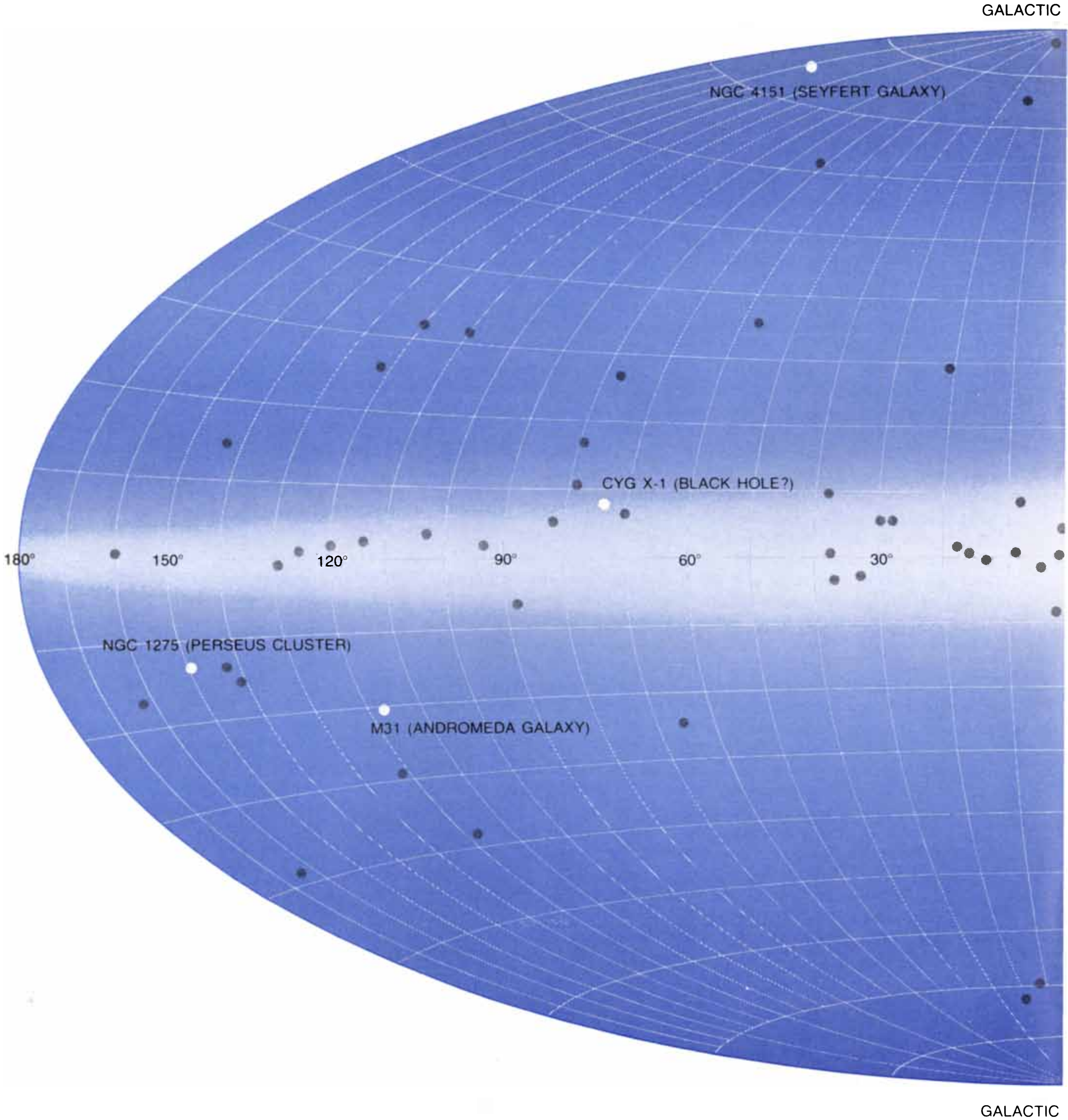
The first X-ray satellite, named *Uhuru* (the Swahili word for "freedom"), was launched in December, 1970, from San Marcos Island off the coast of Kenya. It is still functioning. Its purpose is to survey the sky for new X-ray sources, and it has been quite successful. The latest list, numbering 116 sources, has been derived from only a partial analysis of the available data. Interestingly enough, a fourth of the new sources lie away from the central plane of the galaxy, which suggests that they are extragalactic objects. With these results and the prospect of more discoveries to come, the field of X-ray astronomy has crossed the threshold to maturity.

An astronomer studying an optical object traditionally undertakes to measure such properties as its location on the celestial sphere, its angular size, the distribution of energy across its spectrum, its intensity, the degree to which its light is polarized and whether or not these properties vary with time. The X-ray astronomer does much the same thing, but there is one important difference. In studying the X-ray sky he must differentiate between those properties of sources that can be measured directly and those

X-RAY SOURCE GX9 +1 is represented by the computer-generated display on the opposite page, which is based on data gathered by means of detection device aboard a rocket launched by a group at the Massachusetts Institute of Technology. This display, known as a correlation map, is used to locate strong X-ray sources very precisely.

that can only be inferred from the direct measurements. First he must detect the X rays. The workhorse of X-ray detectors is the proportional counter. It yields an electrical signal whose amplitude, or strength, is proportional to the energy of an incident X-ray photon. Such detectors are usually filled with a gas in which a strong electric field is maintained. When an X ray passes through a thin metal or plastic window in the detector wall, it collides with an atom of the gas and initiates a cascade of electrons and positive ions. The motion of the positive ions in the electric field provides the signal for the electronic circuitry that identifies the event and registers it as an X ray. Non-X-ray events caused by charged particles and gamma rays must be recognized and rejected by the detector and its associated electronic logic.

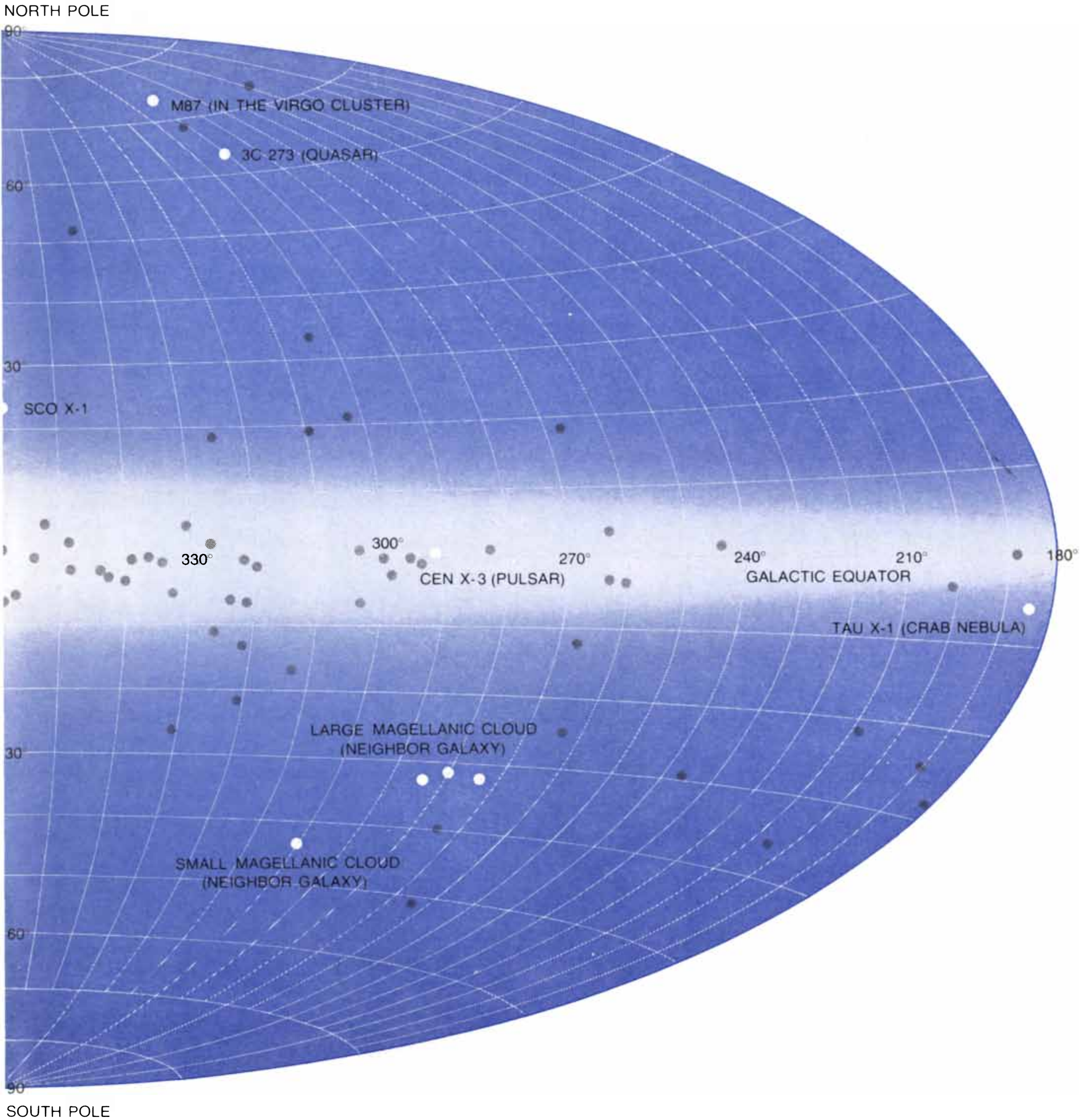
Just as X rays can interact with the gas atoms of the detector, so too can they interact with the atmosphere of the earth. To overcome atmospheric absorption X-ray detectors must be carried aloft by high-altitude balloons and rockets or, better yet, be placed in orbit on satellites. Absorption can also take place in



POSITIONS OF 116 X-RAY OBJECTS are indicated on this map of the galaxy. The locations of the sources have been taken from the 1971 catalogue compiled from data obtained by the X-ray satellite *Uhuru*. Thirteen of the strongest X-ray objects are identified the interstellar gas, and indeed in the X-ray source itself. Such absorption effects can be used to study the nature of X-ray sources and the celestial material along the line of sight.

The ingenuity of the electronics engineer provides the circuitry that encodes into numbers the sequence of the signals from the proportional counter. The numbers are telemetered to the ground, where they are organized in a spectrum to display the relative flux of X-ray photons in consecutive energy intervals. Such spectra, called pulse-height distributions, are exploited in a variety of ways to characterize X-ray sources. The sources can be categorized as being either "hard" or "soft." If the curve of the flux plotted against the energy is steep, indicating a relative paucity of high-energy photons, the source is called soft. If the curve is flat, indicating a strong flux of high-energy photons, the source is called hard. In general the spectrum as measured with a proportional counter is a broad, slowly varying function of the energy called a continuum.

Sometimes restrictions on telemetry,

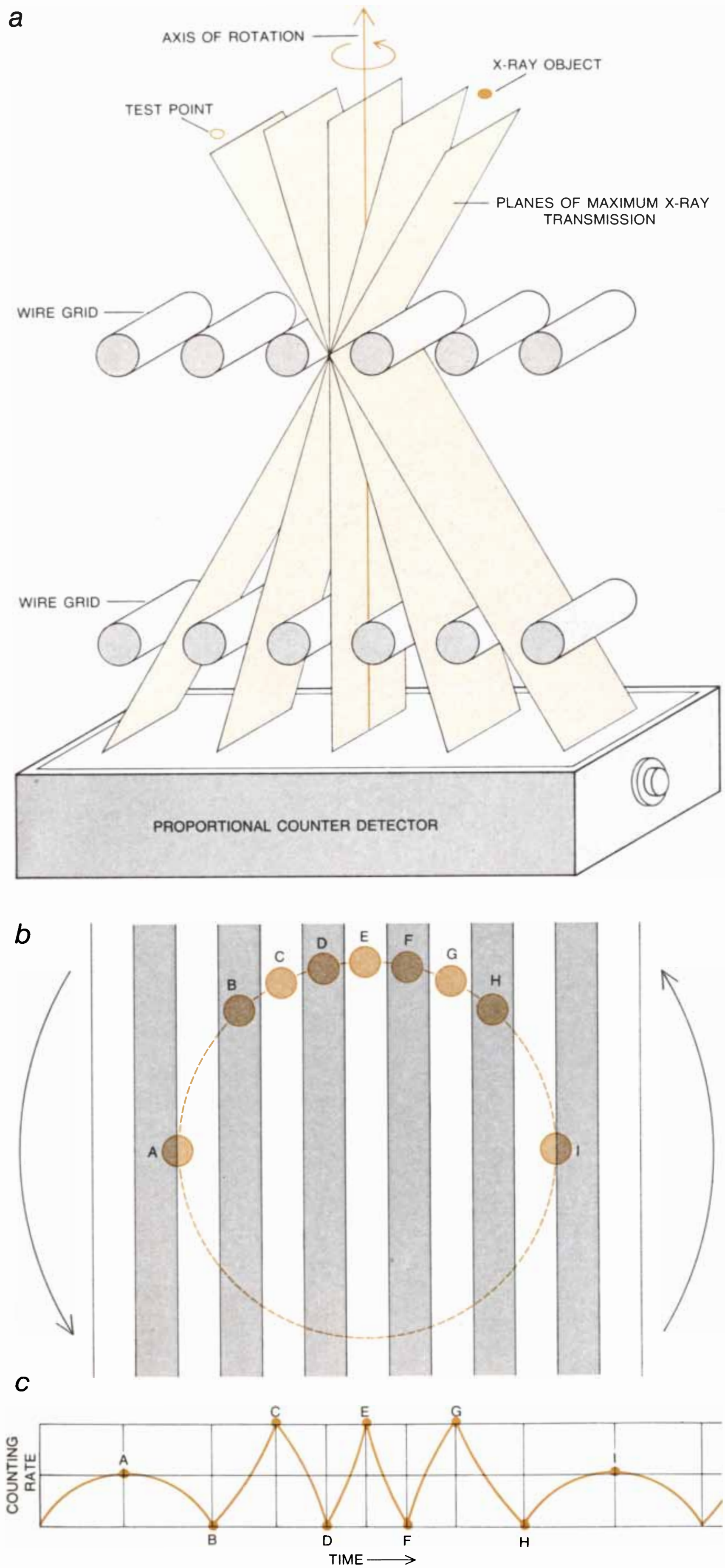


with their visual counterparts. Three-quarters of the objects cluster along the central plane of the galaxy, suggesting that they lie within it. The remaining sources are scattered widely above and below the galactic plane, indicating that they may be outside the galaxy.

weight or power necessitate cruder energy-measuring techniques. Under these circumstances any X-ray signal, regardless of its amplitude, is recorded in the same way as any other signal, and the proportional property of the X-ray detector that relates the amplitude of the response to the strength of the X-ray signal is lost. With the appropriate choice of the detector window material and the gas filling the detector, the signals can be made to correspond to photon energies of a certain range. A sandwich of three or four detectors can be constructed so that low-energy photons are registered (and absorbed) in the first detector, higher-energy photons are registered in the second, and so on. The data so obtained are somewhat analogous to the three-color plots of the optical astronomer.

How are the X-ray sources located on the celestial sphere? Most X-ray astronomers employ essentially the same technique. They simply restrict the field of view of a detector that is otherwise sensitive to X rays arriving from many directions. In order to understand how this works, think of looking at the sky through a ground-glass screen. Obviously it would be difficult to orient the screen so that it is perpendicular to the line of sight toward a particular star, let alone to distinguish one star among many. Suppose, however, a narrow tube is placed in front of the screen. Now the screen will be illuminated only when the tube points essentially at the star, and stars lying outside the angular aperture of the tube will not contribute to the illumination.

Instead of a tube the majority of in-



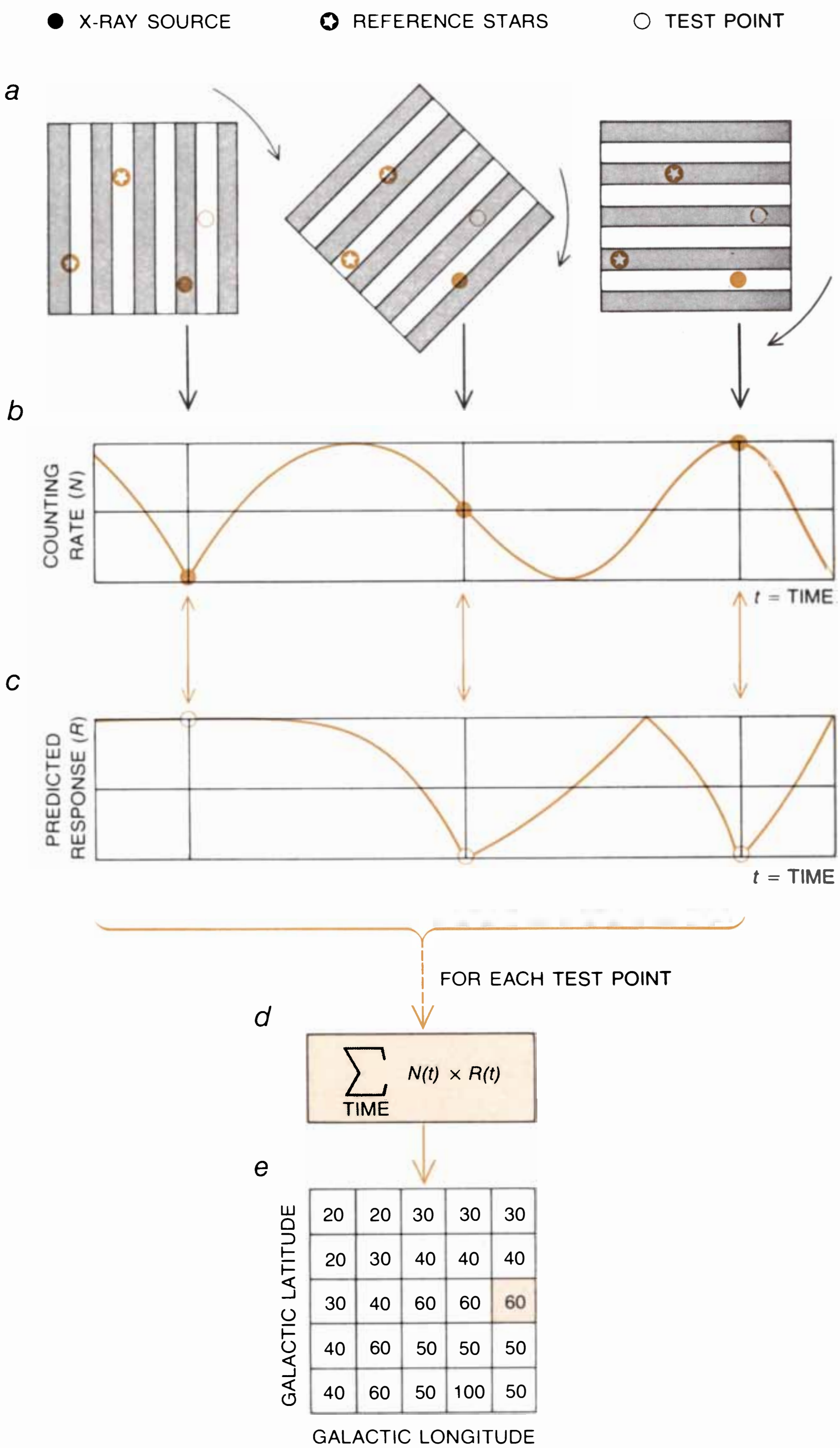
ROTATING MODULATION COLLIMATOR (*a*) can detect the position of an X-ray object to an accuracy of about 10 seconds of arc. Two parallel wire grids define a set of planes of maximum transmission that intersect the celestial sphere in a series of great circles. When the collimator is rotated around the pointing axis (*b*), the strength of signal from each X-ray source in the field of view is modulated as it is alternately eclipsed by the wire grids and exposed to the view of the detector. The light bands represent the transmission planes projected against the sky; the dark bands are areas of little or no transmission. The pattern of the modulation (*c*) is characteristic both of the object's angle of displacement from the pointing axis and of the angle of rotation of this axis measured with respect to some absolute reference angle. Such a transmission pattern is calculated for each test point in the rotating collimator's field of view, allowing the location of X-ray object on sky to be displayed on a computer-generated map.

struments used to search for X-ray sources have a series of parallel metal slats, rather like an open venetian blind, in front of the detector. Since the slats are parallel, the angular response of a system with multiple slats is identical with the response of a system with only two slats. Each pair projects exactly the same narrow viewing band on the sky. As the rocket rotates, the slat-collimator detector system scans the sky in a direction perpendicular to the long dimension of the slats. The detector will therefore view the sky through one band at a time. When a viewing band passes over a source, the counting rate in the detector will increase above the background level.

This procedure does not give the location of a source within a band, nor does it resolve the ambiguity created by several sources lying within the same band. Nevertheless, by rotating the collimated detector around its viewing axis to a new orientation and scanning across a band known to contain a source, it is possible to bracket the exact location of the source to about a tenth of a degree. The crossed scan also helps to reduce the multiple-source ambiguity. It can even give information about the angular size of an extended X-ray object, since such a source can illuminate several contiguous bands in both scan directions. Unless there is some compelling reason to associate a newly discovered X-ray source with a nearby visible object (such as a supernova remnant located within the bracketed area), the positional accuracy obtained with the slat collimator is insufficient for optical identification. Fortunately the accuracy can be markedly improved by a modification of the slat collimator called the rotating modulation collimator.

The slats are replaced by a parallel grid of closely spaced wires. A second such grid positioned above the first defines bands of transmission on the celestial sphere that are separated by just a few minutes of arc. The detector and the grid system are rotated continuously around an axis perpendicular to the grid; the flux from a source is thereby modulated in time. A source lying right on the axis of rotation gives no modulation at all; a source lying off the axis gives a time-varying response that can be analyzed to obtain both the azimuth and the zenith angles of the source with respect to the axis of rotation. If the orientation of the detector with respect to the celestial sphere is accurately known, one can obtain the position of any X-ray source to about 10 seconds of arc.

In addition to the accurate location



DATA FROM ROTATING COLLIMATOR is analyzed indirectly since the collimator does not take photographs of the X-ray sky. As the collimator rotates (*a*) its view of the sky represented by transmission bands changes. The counting rate *N* from the X-ray source varies with time (*b*) as the collimator rotates; the three dots indicate the particular counting rate associated with each of the three sky views. A position for the source is guessed (*open circle*), and the response *R* of the collimator is computed for that test point (*c*). This response is a measure of the counting rate the system would have exhibited if the X-ray source had been located at the test point. At every instant of time corresponding counting rates on the two graphs are multiplied, and the sum of all the resultant products is taken (*d*). This sum, or "correlation function," is related to the probability that the source is actually at the test point. Since the location of the test point with respect to the reference stars is known, the sum can be entered on a map of the sky (*e*). This correlation process is repeated for many test points. A scale of lighter and darker shades of gray can be assigned to the numbered boxes to produce a computer-generated picture on a cathode ray tube.

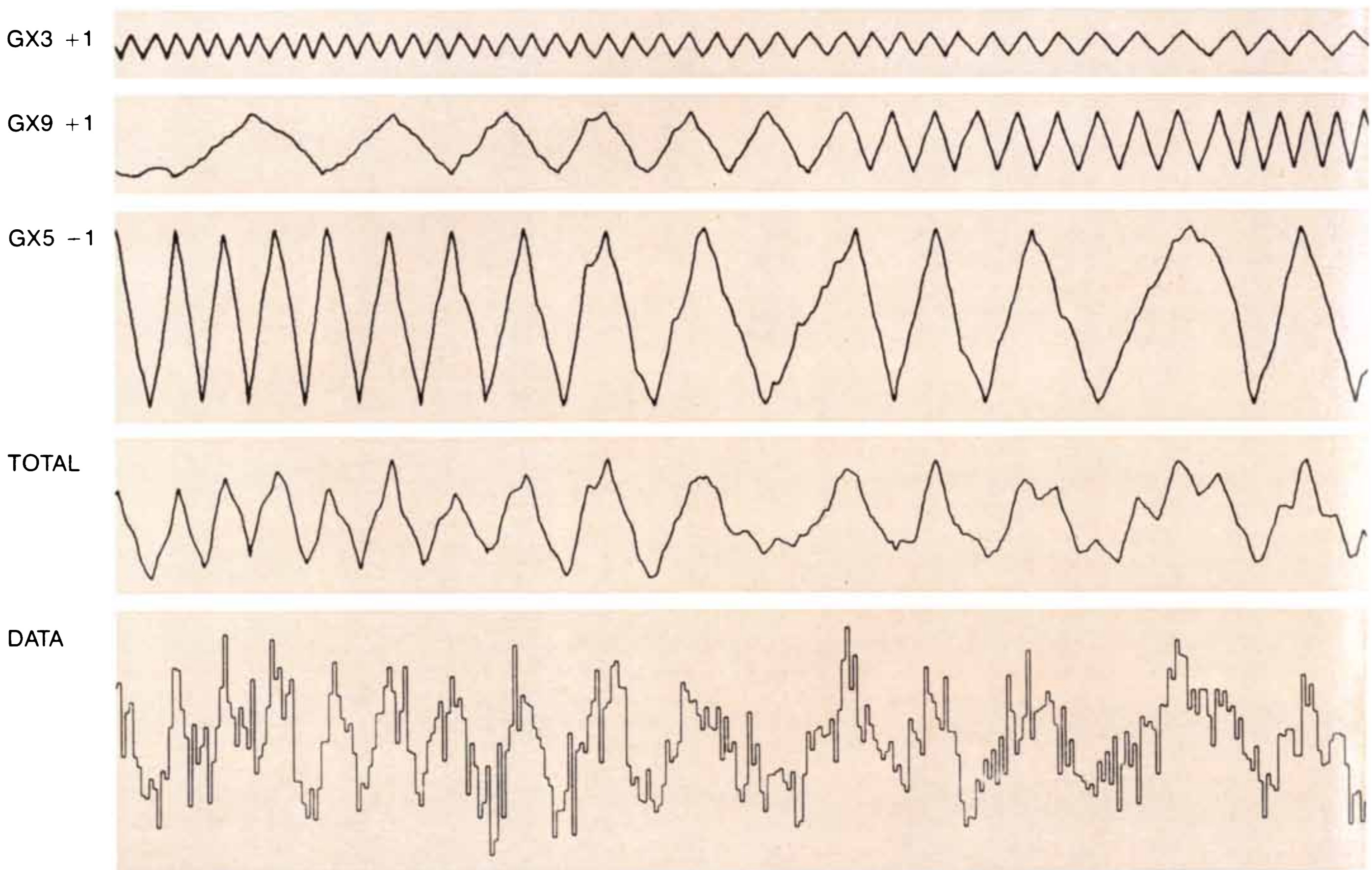
of sources on the celestial sphere, the rotating modulation collimator yields longer viewing times, an improved signal-to-noise ratio and unambiguous identification of multiple sources. A drawback of the system is that most extended sources cannot be detected, since they span several transmission bands and give no modulation when the detector rotates.

One might ask why X-ray telescopes analogous to light telescopes are not built instead of the seemingly complex collimator systems. As a matter of fact, investigators at American Science and Engineering have developed and flown focusing X-ray telescopes to obtain X-ray pictures of the sun [see "X-Ray Stars," by Riccardo Giacconi; SCIENTIFIC AMERICAN, December, 1967]. The best of these telescopes has an angular resolution of about one second of arc. Unfortunately the large physical size that is needed to enhance the sensitivity of such instruments to distant X-ray sources restricts their use to large-satellite missions.

In astrophysics the principal yardstick for measuring theoretical models of stars and other objects is spectroscopic analysis. X-ray spectroscopy is a commonplace laboratory procedure. Application of the same technique to the radiation reaching a spacecraft from an X-ray star is a major challenge to the astronomer. Slice by slice the broad continuum must be examined with high resolution so that the presence of characteristic emission or absorption features can be detected. The low counting rates of most instruments limit this type of investigation to satellite experiments, where long observing times and good pointing accuracy ensure that enough X rays can be detected to free the interpretation of the data from statistical uncertainties.

High resolution is obtained with certain crystals whose properties of reflection disperse the X rays into a spectrum in much the same way that a diffraction grating disperses light. Unlike a grating, however, the crystal reflects only a narrow range of X-ray wavelengths for each angular setting with respect to the incoming beam; a scanning sequence over many incident angles is required to map out an entire spectrum. Several experimental approaches are being taken to get around the problem of low counting rates in future spacecraft flights. Basically they couple a crystal spectrometer to a device that will concentrate the X rays into a beam. Compact detectors can then be used for the beams of incoming X rays. Sensitivity to weak spectral features is thereby enhanced at the cost of complexity.

It is an interesting fact that such crystal analyzers are sensitive to the plane of polarization of the incoming X-ray beam when the beam's angle of incidence to the crystal surface is 45 degrees. In general this angle corresponds to a small band of wavelengths in the continuum, with the center frequency of the band set by the choice of the crystal. Some models for cosmic X-ray production are based on the acceleration of electrons in a magnetic field, and such models predict that the source will emit X rays that are preferentially plane-polarized. The plane of polarization can be found by rotating the crystal around the direction of the incoming beam while maintaining the beam's incident angle of 45 degrees. Crystals that have high reflectivity rath-



PORTION OF DATA (*bottom*) **from the rotating modulation collimator is shown along with the transmission function for each of three test points** (*top three traces*). **Analysis consists in determining such a function for each point in the field of view and correlating it with the data in order to build up a correlation map by computer. When the predicted responses for the correct test points**

er than high resolution are particularly desirable for these studies. Devices that are still more sensitive to polarization effects can be constructed by bending the crystals into optical elements that will focus the X rays on detectors of smaller area.

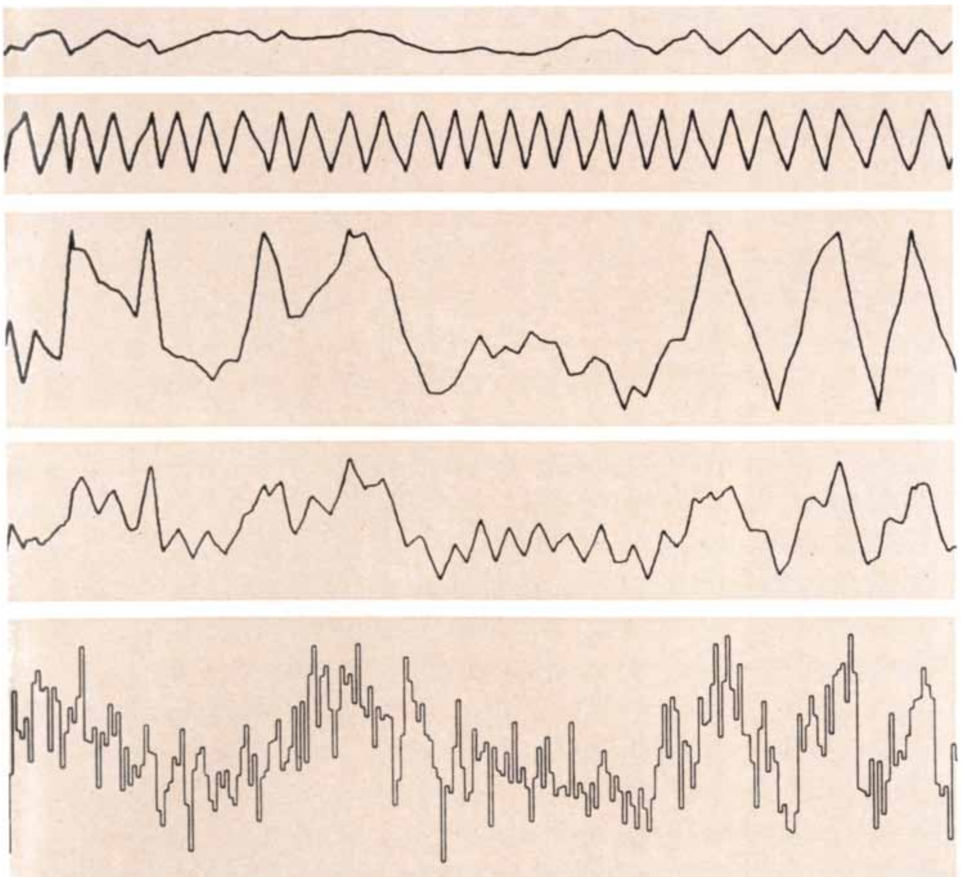
A suitable model of the X-ray source is necessary for interpreting experimental results beyond the directly measurable quantities. Some models are relatively simple and can be derived from a dimensional analysis or an intuitive approach. Others are gigantic efforts of theoretical physics where X-ray observations play only a partial role in the chain of argument. In the long run the usefulness of the model depends on the skill of the astronomer.

Strong X-ray sources are plainly the best candidates for detailed studies. Adequate statistics can be accumulated even with the brief observing times allowed during a rocket flight. Most sources of this kind are characterized by a continuous distribution of intensity extending over a wide range of energies that slowly varies with time. The shape of this continuum can be tested against the shape predicted by various models in their attempt to explain the origin of the X rays. The models are characterized by such parameters as the temperature of the electron distribution or the strength of the magnetic field. The interpretation of the experimental results is rarely unambiguous, and since many of the sources exhibit large temporal effects the models are not yet complete.

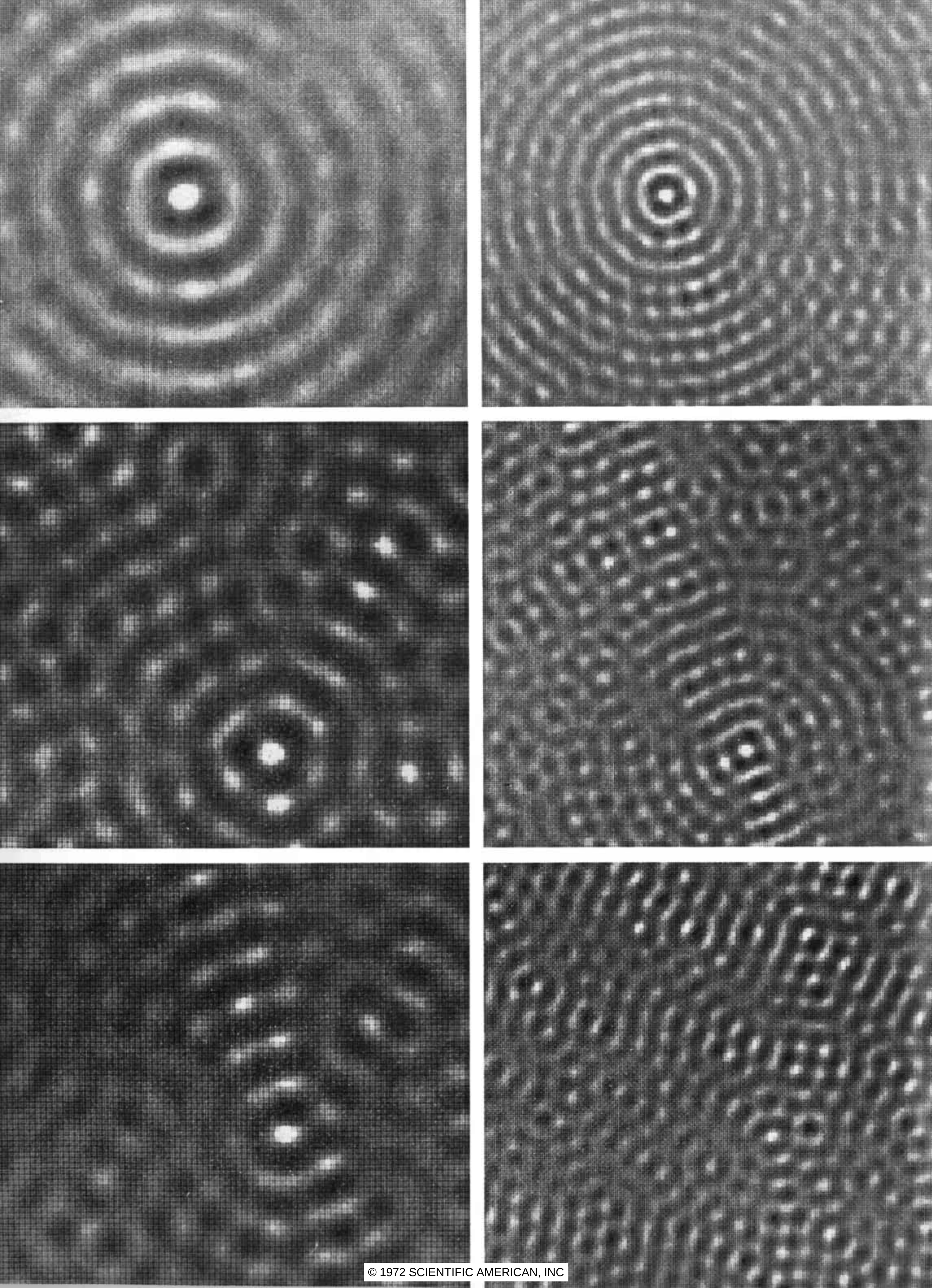
The highly luminous source Sco X-1 has been studied intensively. It is almost never the same from one observation to the next: there are erratic long-term and short-term variations in both its X-ray and optical emissions. Sudden changes in X-ray flux have also been reported. In addition to the changing luminosity the parameters describing the shape of the spectrum are different from one observation to the next. In spite of these continuous changes the spectral observations do permit some generalization. Whenever enough data to construct a spectrum are acquired in a single observation, that spectrum seems to have a shape that is an exponential function of the ratio of the X-ray energy to a characteristic temperature, or energy. A characteristic temperature is a convenient parameter and is part of a simple thermal model that predicts the exponential spectrum.

The model assumes that the emitting region of the X-ray star is a hot, moderately thin plasma, that is, a dilute ionized gas. In this case the gas is mostly hydrogen. If the temperature is sufficiently high, the electrons and protons in the plasma move quite rapidly with frequent close encounters. The electric Coulomb force between the particles is very effective at a short range and accelerates the particles, causing them to be deflected through large angles. This combination of conditions is all that is needed to produce the X rays, since accelerated charges are a source of electromagnetic radiation. A compact name for the process is "thermal bremsstrahlung." In a dilute plasma many layers of the source contribute to the X-ray emission, and the shape of the spectrum is exponential. A somewhat denser plasma will absorb and scatter lower-energy X rays in its outer layers, distorting the exponential spectrum and causing it to turn down slightly at very low energies. The fact that an exponential fit is not exact should not be too disturbing. It is unlikely that the source is ideally thin or that any single characteristic temperature will accurately describe the entire emitting region of the object. The interesting point is that the observed characteristic temperature varies from its normal value of 50 million degrees Kelvin. It has been speculated that the primary source of energy driving the thermal processes could be gravitational energy that is released by the plasma streaming between the partners of a close binary system.

Tau X-1, the X-ray source in the Crab Nebula, presents an entirely different picture. It is known to be extended over a large portion of the diffuse region of the nebula. Between 10 and 15 percent of the X rays come directly from a pulsar, the core of a supernova remnant. The observed spectrum can be theoretically described by a power law of the X-ray energy extending from the radio and optical bands into the high-energy X-ray and gamma-ray region. The spectrum is predicted by a model that also depends on an acceleration process to produce X rays; here, however, the process is magnetic rather than thermal. Highly energetic electrons circulating in a weak magnetic field form a giant synchrotron, a magnetic accelerator that produces radiation corresponding to the required power-law spectrum. In parallel with the



are totaled (*fourth trace*), they compare very well with the data. In the labeling of the sources the letters *GX* refer to the galactic center and the numbers refer to galactic longitude and latitude. Correlation maps of the three sources are shown on pages 26 and 34.

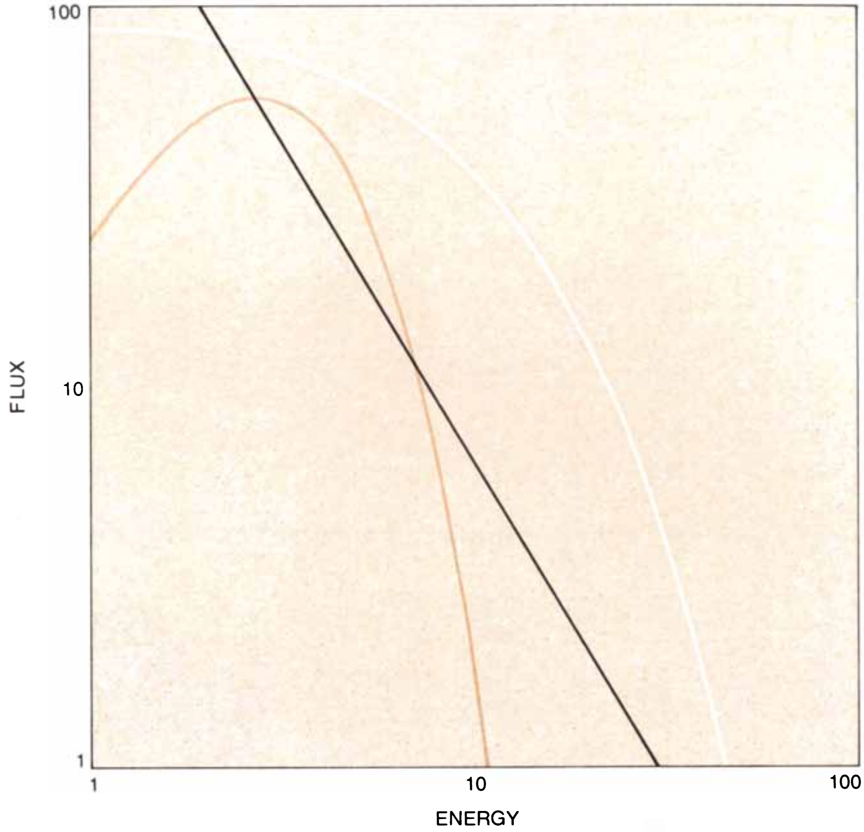


thermal model the energy of the electrons is rapidly degraded as they emit radiation, and a continuous source of energetic electrons is necessary in order to sustain the emission process.

Again the energy source seems to be the pulsar. One model, which proposes that the pulsar is a rapidly spinning neutron star, supplies the frequency of the observed pulsations and meets the energy requirement nicely. Immense quantities of energy are stored in the rapid rotation of the neutron star. Just as a spinning top slows down through the frictional contact between its tip and the floor, so does the pulsar. The "friction" is a complex interaction of the star with its surrounding electromagnetic environment. The net result is that some of the rotational energy is transferred to electrons that stream out of the object, interact with the magnetic field and radiate. Observations confirm that the time between pulsations is increasing at a rate that is sufficient to supply the necessary X-ray energy from the loss of rotational energy.

Distances to X-ray objects can be measured directly if the source is easily identifiable with an optical or radio object whose distance is known. Extragalactic sources and the more spectacular objects within the galaxy are the most promising candidates for this interpretation. There are other methods for determining the distances to unidentified objects within the galaxy. Interstellar gas tends to be somewhat opaque to low-energy X rays. A sufficient amount of gas between the source and the observer will show up noticeably as a cutoff of the low-energy region of the spectrum. Of course, the effects of local gas clouds surrounding the source must also be considered. Moreover, the composition and density of the gas must be measured reliably in order to accurately compute its effect on the continuum spectrum. The distances to several sources, notably those lying in the direction of the galactic center, have been estimated in this way.

Sharp low-energy cutoffs have been noted in the spectra of several objects. One of the most striking is the case of Cyg X-3 in the constellation Cygnus. If this source, like Sco X-1, is patterned on the hot, thin plasma model, the downturn in the spectrum at low energies may be attributed to the absorption mechanisms discussed above.



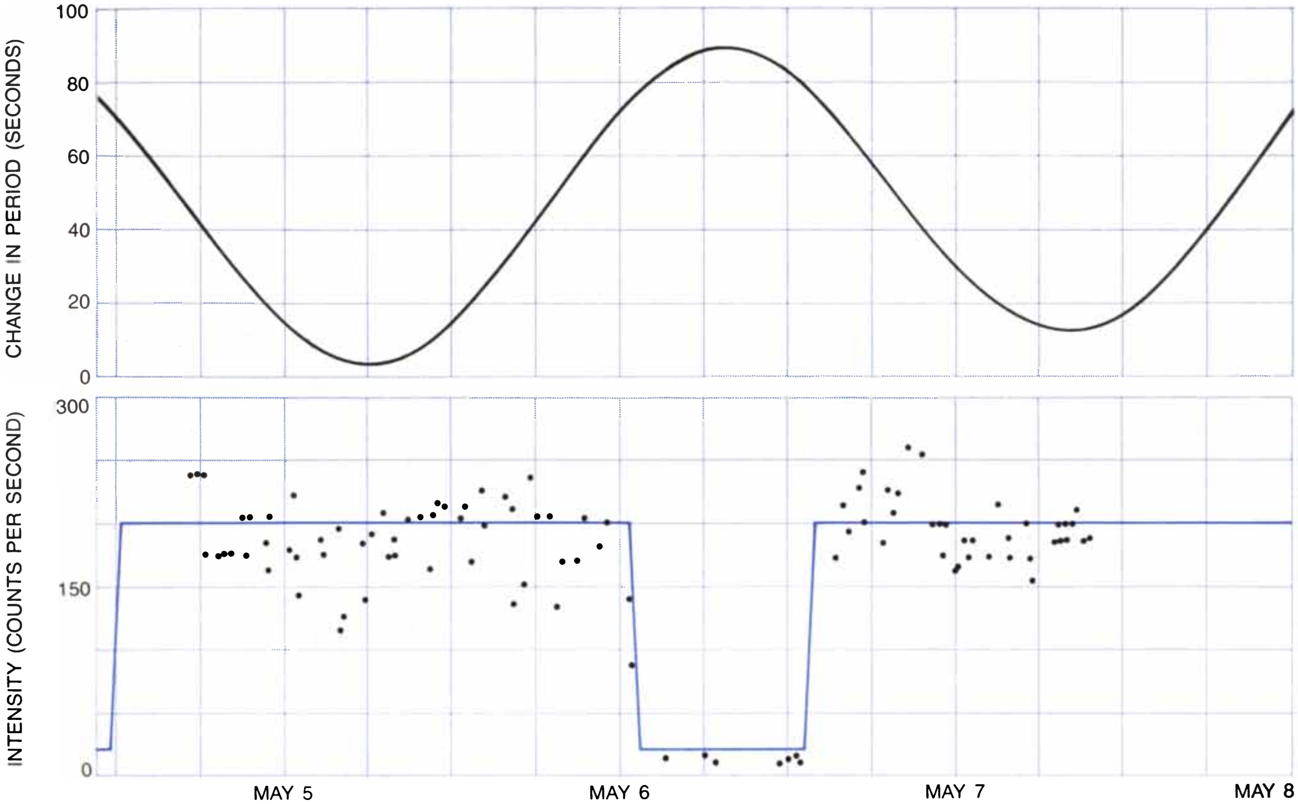
SHAPE OF SPECTRUM differs depending on the mechanism within the object producing the X rays. The horizontal and vertical scales are in arbitrary units. Black-body spectrum (*color*) is produced by a source surrounded by a dense plasma (ionized gas) that absorbs radiation from all of the object except its outermost shell. Synchrotron spectrum (*black*) results from highly energetic electrons spiraling around lines of force in a magnetic field. "Thermal bremsstrahlung" spectrum (*white*) results from the motion of electrons in a hot, moderately thin plasma. The spectra of actual X-ray sources would be modified from these ideal curves because some of the radiation would be absorbed by the interstellar medium.

THREE REGIONS OF X-RAY SKY are displayed in the six panels on the opposite page. The collimator's resolution in the left-hand panels is only eight minutes of arc, whereas the resolution of the right-hand panels is four minutes of arc. The panels show X-ray sources at the positions GX5 −1 (*top*), GX3 +1 (*middle*) and GX17 +2 (*bottom*). The traces of the transmission functions for sources GX5 −1 and GX3 +1 are shown in the illustration on pages 32 and 33.

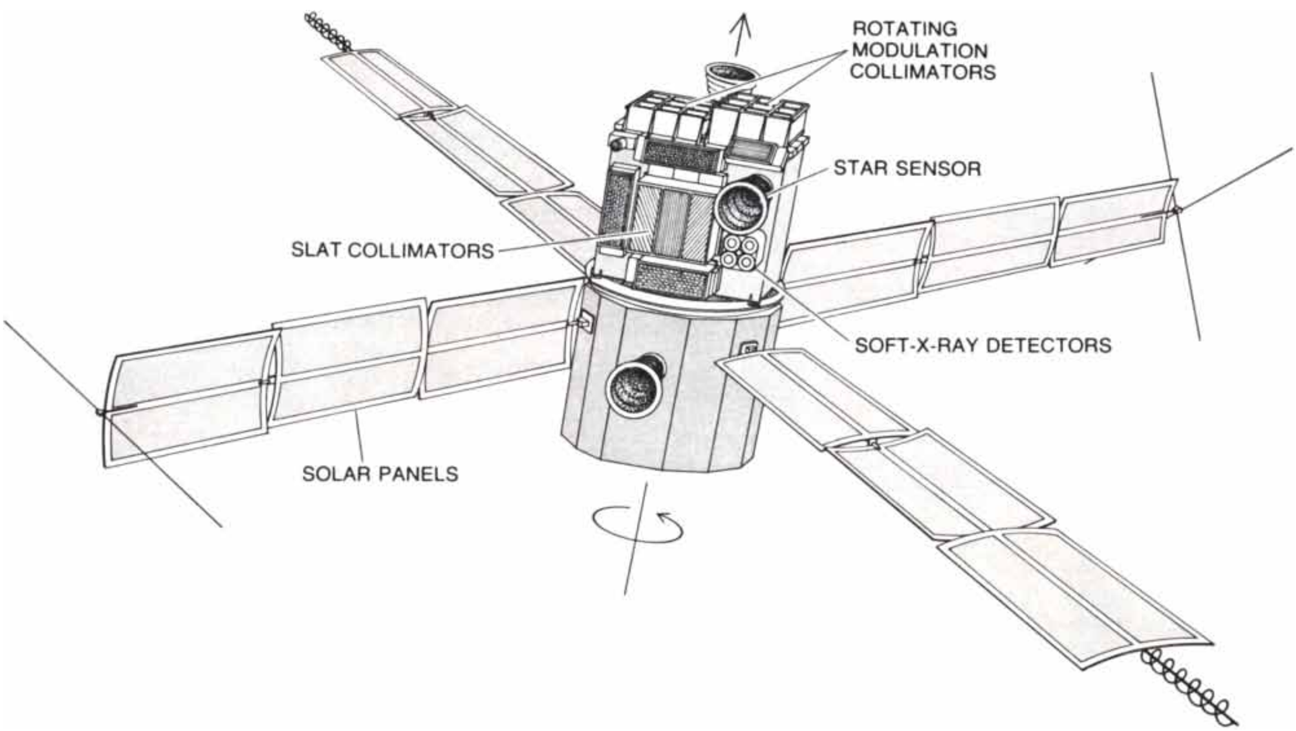
There is, however, another thermal model for X-ray emission that has a built-in downturn. When the plasma is very dense, it absorbs radiation from all parts of the object except its outermost shell. The object behaves in much the same way as a theoretical "black body" and can be described by the same formulas. A black body is defined as a body that emits the maximum amount of heat radiation, and the distribution of its energy with respect to wavelength has a special shape called the Planck distribution [*see illustration above*]. The description of the star's spectrum therefore changes from an exponential distribution to a Planck distribution.

Either the thin-plasma-with-absorption model or the black-body model fits the data from Cyg X-3 quite well. Additional parameters must be found in order to make a unique choice. Of the many sources found with downturns in their spectra at least some must be thin-plasma sources. For them the number of hydrogen atoms along the line of sight from the earth to the source can be estimated from the characteristics of the downturn. There are several times 10^{22} hydrogen atoms per square centimeter along the line of sight toward the galactic center. Estimates of the interstellar distribution of hydrogen can be obtained from data provided by radio astronomers. With this information we can infer the distance to the X-ray sources.

The diffuse X-ray background is a difficult subject to study. At present the observations indicate that this background is fairly isotropic (uniform in all directions) in the energy range between 1,000 and 100,000 electron volts and follows a rather well-defined power-law spectrum. The fact that the radiation is isotropic suggests that these X rays origi-



CEN X-3 IS AN ECLIPSING BINARY, as is shown by periodic variation in the counting rate from the *Uhuru* satellite. The X-ray source pulsates with a period of 4.8 seconds. Since the pulsation rate is constant it is possible to predict the time at which each pulse should occur as measured from some reference time. It was found that the arrival time of the pulses varies sinusoidally every 2.1 days (*top*). This variation is due to the Doppler effect as the X-ray source orbits around its companion body. The intensity of X radiation also varies every 2.1 days as the X-ray source is eclipsed by the companion (*bottom*), which lies in the plane of the line of sight.



SAS-C SATELLITE will be launched in late 1973 to study the X-ray sky in more detail. This version of the satellite has its solar panels deployed. The three "eye stalks" are sun shields to protect the star sensors that determine the attitude of the spacecraft. A cluster of four small tubes on one side of the slat collimators in the center of the satellite scan the galactic equator for soft X rays. Rotating modulation collimators are contained in the two boxes on the very top of the satellite that look out along the spin axis. Other experiments will study time variations of Sco X-1, accurately locate X-ray sources in galaxy and search for extragalactic X-ray sources.

nated outside the galaxy, and two possible origins have been proposed. The first attributes the background to a superposition of unresolved discrete sources. The second suggests that the diffuse X rays may be a product of phenomena in the medium between galaxies. It has been argued that the contribution of all the galaxies in the universe, if they are not very different from ours, is not enough to produce the observed flux. Perhaps a fraction of the other galaxies are much brighter than our own, but then we might see a significant departure from the even distribution of X rays over the background.

Observations at the low-energy end of the spectrum (less than 1,000 electron volts) may resolve these problems. At low energies interstellar absorption becomes important. Careful examination of the data suggests that diffuse X rays do not originate entirely outside our galaxy but are in part produced within the galaxy.

One of the mechanisms for generating diffuse X rays in extragalactic space is thermal bremsstrahlung by extragalactic electrons, requiring an extraordinarily dense, hot intergalactic gas. That requirement can be avoided, however, since it is possible that bremsstrahlung could arise from interactions of the electrons with extremely-high-energy cosmic ray protons. Another proposed mechanism involves inverse Compton scattering, a process wherein an energetic electron interacts with the photons of the universal three-degree-Kelvin blackbody radiation permeating space and transfers enough kinetic energy to make an X ray out of a low-energy photon. This process requires that the abundance of extragalactic electrons be about 1 percent of the abundance of galactic electrons.

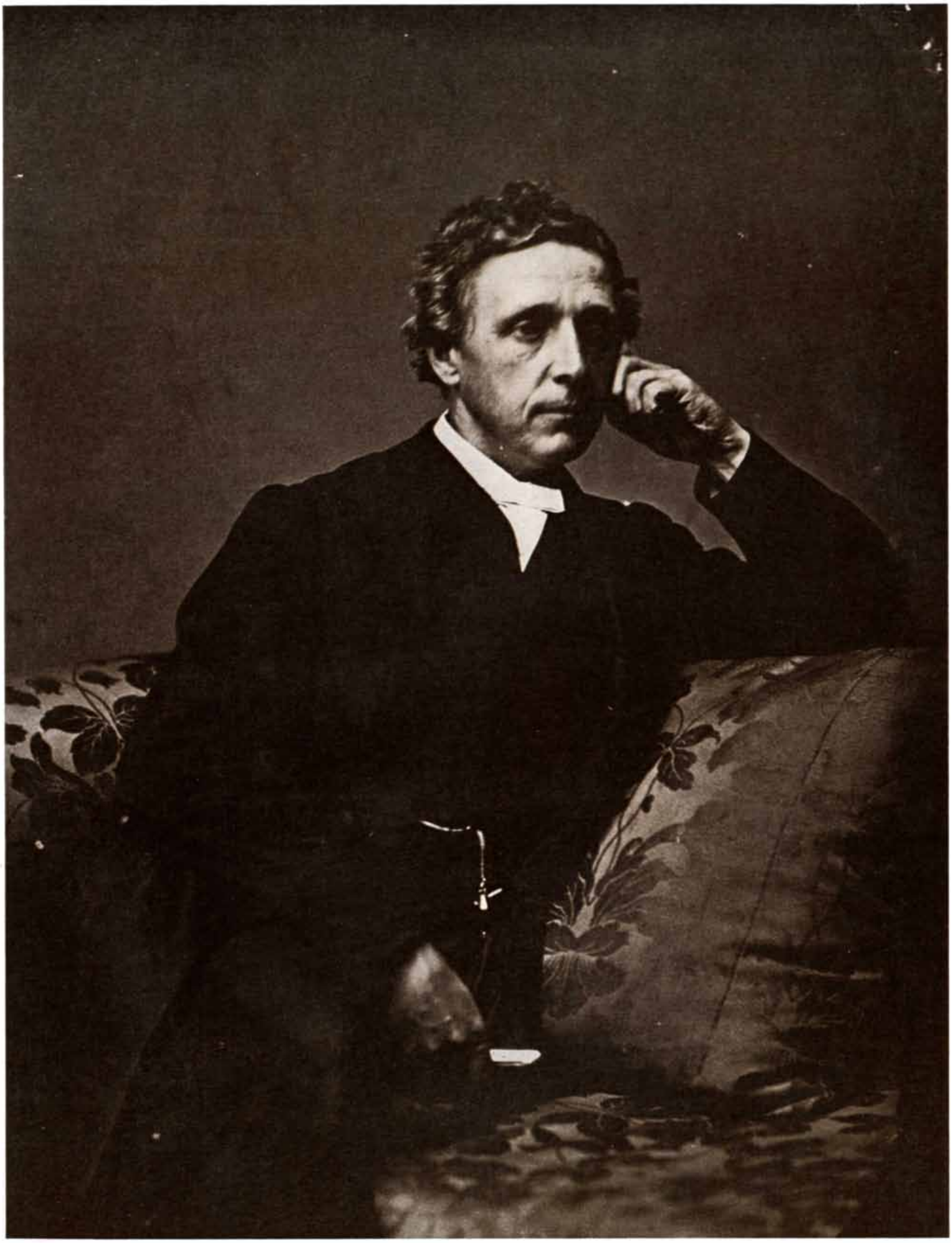
With the exception of the total-reflection focusing X-ray telescope, none of the collimation systems either in use or proposed has the ability to resolve stellar X-ray sources. The size of the X-ray object is either built into the model (as in the neutron-star hypothesis for the X-ray pulsar in the Crab Nebula, where a neutron star has a limiting diameter) or must be inferred from the observations. If, for example, a thermal model fits the spectrum, then it is possible to establish the surface brightness of the source, that is, the flux of radiation leaving each unit area of the object. If the distance to the source is also known, the flux measured at the detector can be used to determine the total flux radiated by the source and therefore its total area. The diameter follows directly from the area. Similar arguments hold for sources where the X rays are produced by magnetic effects. So far these arguments have not worked too well.

The tendency for the spectral output of the X-ray sources to vary with time also plays an important role. Although the pulsar in the Crab Nebula was first discovered at radio frequencies and was later studied visually, most of its energy is emitted in the X-ray range. The neutron-star hypothesis for this pulsar is supported by the excellent agreement between the X-ray, radio and optical results. The process that converts the rotational energy of the neutron star into pulsed electromagnetic radiation, however, is not clearly understood; it does not follow directly from the observation of pulsed X rays. The sinusoidal variation of the short-term periodic behavior of Cen X-3 has been interpreted as indicating that the source is one member of an eclipsing binary system. The observations agree strikingly with this simple model. Evidence for a binary system has also been found for Cyg X-1.

Much remains to be done in the way of understanding the behavior of matter that exists under the conditions prevailing in X-ray stars. It would be extremely valuable to be able to reproduce a representative piece of an X-ray star in the laboratory. In fact, a group at M.I.T. is actually preparing such an experiment in connection with a project on the containment of a plasma for fusion reactors. Large quantities of detailed information should make it possible to construct more accurate models against which the data from X-ray stars can be tested.

If emission lines are eventually detected in the spectrum of an X-ray source, there is much to be learned beyond the knowledge that some particular chemical element is present as a constituent in the object. Characteristic X rays are produced when an inner electron is removed from an atom; the vacancy thus created is filled by another electron at some later time and an X ray characteristic of this process is emitted. The inner electron may be removed during the collision with one of the energetic electrons in the hot plasma of the X-ray source. The electron that fills the vacancy may be one of the electrons that remain bound to the atom in outer shells or it may come from the plasma itself. From the shape of the spectral lines may come such important information as density, collision rate, motion and temperature. We can infer some of the same parameters from the continuum spectrum, but the two sets of data may not agree because different regions of the source can be responsible for each kind of emission. This nonequilibrium behavior is well known from X-ray studies of the sun. Several laboratories are now developing instruments that will be capable of making such detailed spectroscopic analyses of X-ray stars for future spacecraft flights.

At this moment Orbiting Solar Observatory 7 (OSO-7) is in orbit. On it is an M.I.T. experiment designed to survey the X-ray sky in five broad energy ranges spaced in a roughly logarithmic way from just under 1,000 electron volts to about 60,000 electron volts. Coming up during 1973 is the launch of the second of the Small Astronomical Satellite X-ray vehicles (SAS-C). Developed at M.I.T., it is a multifunction spacecraft that is capable of locating sources with extreme accuracy, analyzing broad bands of their spectra, examining their variations with time and studying the wide-angle distribution of the diffuse X-ray background. Similar experiments will be flown in 1974 on a British spacecraft (UK-5) and the Netherlands National Satellite (ANS). In the middle or late 1970's the spacecraft of a new series will make their appearance. The High-Energy Astronomical Observatory (HEAO) series is the largest and scientifically the most ambitious unmanned satellite program yet undertaken. The first two flights have been scheduled as survey missions with large-scale versions of experiments that are now being flown and tested on other vehicles. The third and by far the most exacting mission will have a large X-ray-concentrating telescope aboard. The instrument will be about a meter in diameter and will have a focal length of nearly seven meters; its angular resolution will rival that of many large earthbound telescopes. With other equipment aboard there will be a series of measurements coordinating spectra, time variations and polarization effects. The spacecraft will be maneuvered through the sky observing one X-ray source after another and will be held fixed on each source for a period of days or weeks as necessary. Thanks to the results of these missions we shall no longer be denied a full view of the X-ray sky. From these experiments will come the answers to many of the complex questions about the X-ray sources, and with a little luck many new questions will arise to confront the next generation of X-ray astronomers.



CHARLES LUTWIDGE DODGSON was born on January 27, 1832, and died on January 14, 1898. The logician who as Lewis Carroll wrote *Alice's Adventures in Wonderland* was also an avid amateur photographer, specializing in portraiture. This portrait is from the Gernsheim Collection at the Humanities Research Center of the University of Texas at Austin. The print was made from the original five-by-six-inch glass negative in the Gernsheim Collection. The negative has the number 2439 in Dodgson's handwriting on it.

Lewis Carroll's Lost Book on Logic

It is well known that the author of "Alice" also wrote a book on symbolic logic. A sequel to that book has recently been discovered. It supports the view that his work in logic was highly original

by W. W. Bartley III

"I'm so glad I don't like asparagus," said the Small Girl to a Sympathetic Friend. "Because, if I did, I should have to eat it—and I ca'n't bear it!"

These familiar-sounding words were written by Rev. Charles Lutwidge Dodgson, the Oxford mathematics don known to the world as Lewis Carroll. They are not, however, to be found in any published work of Dodgson's. The Small Girl and her Sympathetic Friend, like Achilles and the Tortoise, the Crocodile and the Liar, the Three Barbers, the Five Liars and the Pork-Chop-eating Logician and Gambler, are among the many characters, some familiar, some new, who sharpen their wits on one another and on us in the recently discovered manuscript and galley proofs of Dodgson's sequel to *Symbolic Logic: Part I, Elementary,* published in 1896. From that title it is apparent that Dodgson planned further publications on the subject, but the manuscript on which he was working disappeared shortly after his death at the age of 65 in January, 1898. No mention of it can be found in any of the detailed listings of Carrolliana published in the past half-century, and the bulk of Dodgson's papers were burned shortly after his death.

I first found a small portion of the missing work in some of Dodgson's remaining papers at Christ Church, Oxford, in 1959. The tantalizing segment was neatly set up in type as a set of galley proofs. For the next 10 years I searched for more of the work in both public and private collections of "Lewis Carroll" writings. Finally in the vast collection of Dodgson manuscripts and letters assembled by Morton N. Cohen, professor of English at the Graduate Center of the City University of New York, I found some photocopies of other portions of the galley proofs. Cohen had obtained them from originals in the library of John H. A. Sparrow, warden of All Souls College, Oxford. Sparrow had received the papers from the late A. S. L. Farquharson, who had edited the posthumously published papers of John Cook Wilson. A professor of logic at Oxford, Wilson had received the galley proofs in the mail from Dodgson himself on November 6, 1896, and had apparently forgotten to return them. Cohen and Roger Lancelyn Green are now editing a definitive edition of Dodgson's correspondence. The hitherto unpublished material presented here is copyrighted by the Estate of Charles Lutwidge Dodgson. I am now preparing a critical edition of the entire work on logic.

Together with odds and ends of correspondence and manuscript that so far have been largely undecipherable, the galley and manuscript pages confirm what some historians of mathematics conjectured about Dodgson on the slim evidence provided by the first volume of *Symbolic Logic* and several papers he had published in the philosophical journal *Mind.* One historian who recognized Dodgson's original contributions to mathematics was Eric Temple Bell, who made this evaluation: "As a mathematical logician, he was far ahead of his British contemporaries.... He had in him the stuff of a great mathematical logician."

A decade after Dodgson's death his work was overshadowed by the revolution in logic resulting from the publication of *Principia Mathematica* by Alfred North Whitehead and Bertrand Russell. The second part of *Symbolic Logic* shows that Dodgson was one of the most interesting technical innovators in the transitional period between traditional Aristotelian school logic and the new logic espoused by Russell. It also confirms that Dodgson was an unrivaled propounder of problems, puzzles and paradoxes. This is the more impressive when one considers that innovative work in fields such as logic is usually done by young men and that Dodgson did most of his work in logic when he was past 60. He worked alone; the only logician with whom he was in regular contact was Wilson, who provided little stimulation. Wilson bitterly opposed the new symbolic logic being developed by Dodgson and others. Later Wilson was incredulous that Russell, whose work he described as "contemptible stuff," could find a publisher.

The nature of the transition from Aristotelian logic to contemporary mathematical logic is sometimes misunderstood. Some people mistakenly think that Aristotelian logic has been proved to be "wrong," and that it has been displaced by contemporary logic in the same way that a new scientific hypothesis may displace an older one. The difference between traditional logic and contemporary logic is of a different order.

Logicians try to formulate "rules of valid argument" that will ensure that from true premises only true conclusions can be drawn. An argument is valid when and only when no counterexample can be produced. A counterexample is produced if one can reason from a set of true premises to a false conclusion by following the stated rules. The objective is to avoid invalid arguments and the rules of inference that allow them.

Consider the following argument, which is easily handled by a syllogism within the framework of Aristotelian logic:

All men are mortal.
All Greeks are men.
∴ All Greeks are mortal.

The symbol ∴, of course, stands for "therefore." This is a valid inference, and the rule is:

All *M* are *X*.
All *G* are *M*.
∴ All *G* are *X*.

Any argument of this form, no matter what is substituted for *M*, *X* and *G*, will be valid.

Medieval Aristotelian logicians codified the valid forms of inference according to their "figure" and "mood." Variations in the positions of the terms of a syllogism (*M*, *X* and *G* in the example) are called differences of figure. Each syllogism also has a mood that is determined by the form of its component propositions. There can be 15, 19 or 24 or more valid forms of inference, depending on what type of codification is used.

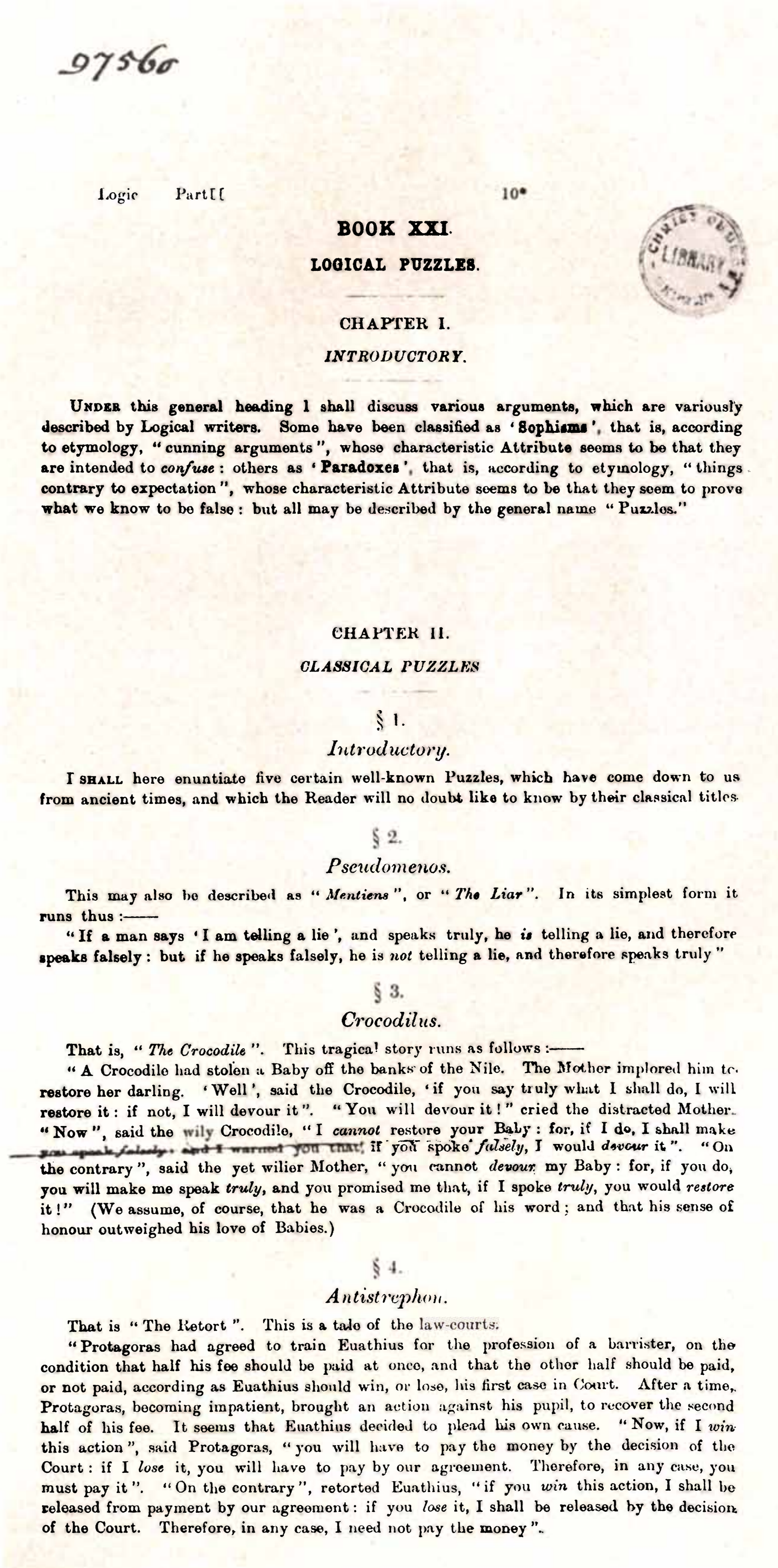
The problem is that there are many valid arguments whose valid rules of inference cannot even be formulated within the framework of traditional Aristotelian logic. For example:

Rebecca is the mother of Jacob.
Jacob is the father of Joseph.
The mother of the father is
the paternal grandmother.
∴ Rebecca is the paternal
grandmother of Joseph.

The argument can be formulated thus in the language of Aristotelian logic:

All *A* are *B*.
All *C* are *D*.
All *E* are *F*.
∴ All *A* are *G*.

Yet once this valid argument is formulated in this way it is quite impossible to state a valid rule of inference for it. A phrase such as "mother of Jacob" is fused into a single term (*B*) and cannot be separated out again. One can easily find other phrases that, when they are substituted for the terms *A* through *G*, will produce a counterexample that leads from a true premise to a false conclusion. The logical structure of the language of Aristotelian categorical propositions is too weak to exhibit how the predicate "mother of Jacob" contains the subject of the second premise and


97560

Logic Part II 10*

BOOK XXI.

LOGICAL PUZZLES.

CHAPTER I.

INTRODUCTORY.

UNDER this general heading I shall discuss various arguments, which are variously described by Logical writers. Some have been classified as '**Sophisms**', that is, according to etymology, "cunning arguments", whose characteristic Attribute seems to be that they are intended to *confuse*: others as '**Paradoxes**', that is, according to etymology, "things contrary to expectation", whose characteristic Attribute seems to be that they seem to prove what we know to be false: but all may be described by the general name "Puzzles."

CHAPTER II.

CLASSICAL PUZZLES

§ 1.

Introductory.

I SHALL here enuntiate five certain well-known Puzzles, which have come down to us from ancient times, and which the Reader will no doubt like to know by their classical titles.

§ 2.

Pseudomenos.

This may also be described as "*Mentiens*", or "*The Liar*". In its simplest form it runs thus:—

"If a man says 'I am telling a lie', and speaks truly, he *is* telling a lie, and therefore speaks falsely: but if he speaks falsely, he is *not* telling a lie, and therefore speaks truly"

§ 3.

Crocodilus.

That is, "*The Crocodile*". This tragical story runs as follows:—

"A Crocodile had stolen a Baby off the banks of the Nile. The Mother implored him to restore her darling. 'Well', said the Crocodile, 'if you say truly what I shall do, I will restore it: if not, I will devour it". "You will devour it!" cried the distracted Mother. "Now", said the wily Crocodile, "I *cannot* restore your Baby: for, if I do, I shall make ~~you speak *falsely*, and I warned you that~~, if you spoke *falsely*, I would *devour* it". "On the contrary", said the yet wilier Mother, "you cannot *devour* my Baby: for, if you do, you will make me speak *truly*, and you promised me that, if I spoke *truly*, you would *restore* it!" (We assume, of course, that he was a Crocodile of his word; and that his sense of honour outweighed his love of Babies.)

§ 4.

Antistrephon.

That is "The Retort". This is a tale of the law-courts.

"Protagoras had agreed to train Euathius for the profession of a barrister, on the condition that half his fee should be paid at once, and that the other half should be paid, or not paid, according as Euathius should win, or lose, his first case in Court. After a time, Protagoras, becoming impatient, brought an action against his pupil, to recover the second half of his fee. It seems that Euathius decided to plead his own cause. "Now, if I *win* this action", said Protagoras, "you will have to pay the money by the decision of the Court: if I *lose* it, you will have to pay by our agreement. Therefore, in any case, you must pay it". "On the contrary", retorted Euathius, "if you *win* this action, I shall be released from payment by our agreement: if you *lose* it, I shall be released by the decision of the Court. Therefore, in any case, I need not pay the money".

GALLEY PROOFS discovered by the author at Christ Church, Oxford, in 1959 were the first clue that major portions of Dodgson's missing second book on symbolic logic might still exist. The number in the upper left corner was penned by Dodgson, who kept a register of all his correspondence. A fold in the galley proof partially obscures a line of type.

part of the subject of the third premise.

With the contemporary logic of relations it is trivially easy to exhibit the valid rule of inference for this example. Let *x*, *y* and *z* stand for Rebecca, Jacob and Joseph, and *M*, *F* and *T* stand for the relations between the individuals: "mother of," "father of" and "paternal grandmother of." Then

'*M x y*'
'*F y z*'
'*M* '*F* = *T*'
∴ '*T x z*'

The rule of inference states that any conclusion of the logical form '*T x z*' is unconditionally deducible from a set of statements of the forms cited above it.

One of the chief aims behind the construction of the logical calculi of contemporary logic is systematically to reduce the valid rules of inference to the smallest number possible. Certainly Dodgson had the same aim. In *Symbolic Logic, Part I* he wrote: "As to Syllogisms, I find that their nineteen forms, with about a score of others which [textbooks] have ignored, can all be arranged under *three* forms, each with a very simple Rule of its own." He regarded Aristotelian logic as "an almost useless machine, for practical purposes, many of the Conclusions being incomplete, and many quite legitimate forms being ignored."

It is apparent that Dodgson was trying to push beyond the traditional forms of valid argumentation. What specific achievements can now be credited to him on the basis of his unpublished work on symbolic logic? First, by 1896 he had developed a mechanical test of validity for a large part of the logic of terms, an achievement usually credited to Leopold Löwenheim. (Löwenheim's decision procedure, developed in 1915, was more thorough than Dodgson's. Both, however, were quite cumbersome. Not until 1922, with the work of Heinrich Behmann, was an easily manageable procedure achieved.)

Second, as early as 1894 Dodgson used truth tables for the solution of specific logic problems. The application of truth tables and matrices did not come into general use until 1920. Third, by 1896 Dodgson had developed "The Method of Trees" for determining the validity of what were by the standards of his English contemporaries highly complicated arguments. The idea was to test whether a conclusion hypothetically assumed to be false, when it was joined with a series of premises assumed

SYMBOL	MEANING
x_1	The subscript 1 asserts the existence of *x*: "Some existing things have the attribute *x*," or more briefly "Some *x* exist."
x'	The *prime* negates a term or statement. If *x* means "new," then x' means "not-new." Thus x' is read "not-*x*."
x_0	The subscript 0 asserts the nonexistence of *x*: "No existing things have the attribute *x*," or more briefly "No *x* exist."
xy_1	"Some *xy* exist." When there are two letters in an expression, it does not matter which stands first; the meaning is the same. The expression also means "Some *x* are *y*" and "Some *y* are *x*."
xy_0	"No *xy* exist," which is equivalent to "No *x* are *y*" and "No *y* are *x*."
$x_1y'_0$	"All *x* are *y*." The subscript 0 asserts the nonexistence of the combination *x* and y'. The expression tells us that some *x*-things exist but that none of them have the attribute y', and in Dodgson's system this is equivalent to the proposition "All *x* are *y*."
†	The *dagger* means "and." Thus xy_1 † xy'_0 means "Some *xy* exist and no xy' exist," or "Some *x* are *y* and no *x* are not-*y*."
⁋	Means "would, if true, prove," or "is derivable from." Thus *x* ⁋ *y* means "*x* proves *y*," or "*y* is derivable from *x*."

METHOD OF SUBSCRIPTS was devised by Dodgson to represent propositions in a shorthand form. Each subscript takes effect at the beginning of the expression it follows, but the prime negates only the term to which it is appended. Dodgson took propositions beginning with "All" ("All *x* are *y*") as being equivalent to two propositions: "There are some *x*" and "No *x* are not-*y*." When converting the "All" propositions into his subscript form, the predicate had to be negated. Thus "All *x* are *y*" is written "$x_1\ y'_0$" and read either as "No *x* are not-*y*" and "Some *x* are *y*" or its equivalent "All *x* are *y*." Similarly, "All *y* are not-*x*" becomes "y_1x_0" ("No *y* are *x*" and "Some *y* are not-*x*"), and the expression "$x'_1y'_0$" is read "All not-*x* are *y*." In translating an "All" statement from one form to the other the predicate (last term) always changes from positive to negative (*x* to x') or from negative to positive.

Consider the premises: xm_0 † ym'_0

Since m' negates *m*, both terms can be eliminated:

$x\not{m}_0$ † $y\not{m}'_0$

What remains can be considered a single expression: xy_0

∴ xm_0 † ym'_0 ⁋ xy_0

For multiple premises in what Carroll calls the first figure, the elimination process is repeated until nothing more can be eliminated. What then remains are the terms that will appear in the conclusion. For example, if the starting premises are:

Data 1.	2.	3.	4.
$a'_1c'_0$	$a_1e'_0$	$c_1b'_0$	d_1b_0

Combine Data 1 and 2, and eliminate $a'a$.
Combine Data 1 and 3, and eliminate $c'c$.
Combine Data 3 and 4, and eliminate $b'b$.

What remains is: $d_1e'_0$. This is the conclusion.

∴ $a'_1c'_0$ † $a_1e'_0$ † $c_1b'_0$ † d_1b_0 ⁋ $d_1e'_0$

DERIVING CONCLUSIONS from premises that are stated in Dodgson's subscript shorthand often involves eliminating terms that negate each other (*m* and m', for example, since m' means not-*m*). Dodgson preferred a method of underscoring the terms to be eliminated rather than the method depicted here, but the results are the same. The solution for the upper problem can be read: "No *x* are *m* and no *y* are not-*m* would, if true, prove no *x* are *y*." In the lower problem, premises can be combined in any order, as long as all are used.

to be true, would lead to a contradiction or an absurdity. His procedure bears a striking resemblance to the trees frequently employed by contemporary logicians, most of which stem from the method of "Semantic Tableaux" devised in 1955 by Evert Willem Beth.

Although Dodgson never attained today's standards of rigor, his anticipation of modern developments in logic is sufficient evidence of his originality. Yet in spite of the anti-Aristotelian character of much of his work, Dodgson did remain belligerently Aristotelian on one issue: the "existential import" of universal propositions. He held to the view that an "All" statement, such as "All men are mortal," was equivalent to two statements: "No men are not-mortal" and "Some men are mortal." Since every "All" statement contains a "Some" statement, all "All" statements assert the real existence of their subjects.

By the middle of the 19th century logicians, notably George Boole, had begun to deny that "All" statements necessarily asserted the existence of their subjects. The Boolean interpretation is almost universally accepted by mathematical logicians today (although it was challenged by Richard B. Angell of Wayne State University in 1964). Thus in the view of most of today's logicians Dodgson's belief in the "existential import" of universal propositions seriously flaws his work. (Nevertheless, Dodgson's decision techniques and formalism can be interpreted so that one usually gets Boolean results rather than Aristotelian ones.)

Although Dodgson's technical innovations are historically interesting, the most fascinating passages of his unpublished writings are those devoted to paradoxes and puzzles. One of them, "A Logical Paradox," was published in *Mind* in 1894 and is still the subject of bewildered controversy among logicians today. The problem involves a barbershop with three barbers who can leave the shop only under certain conditions, which are stated as premises [*see box on opposite page*]. Valid argument from these two premises, however, leads to contradictory conclusions. Dodgson described the barbershop paradox as an "ornamental presentation" of a dispute between Wilson and himself that began in 1893. This was continued for more than a year in a vast output of correspondence, much of which is preserved and which eventually will be published, and also in a series of manuscripts, several of which were privately published by Dodgson. In my critical edition of *Symbolic Logic* I shall reproduce all existing versions of Dodgson's treatment of the paradox. Dodgson indicates in his notes and correspondence that he intended to incorporate the text of "A Logical Puzzle" into the second volume of *Symbolic Logic*.

As the simplest possible example of this Method, let us take the original typical Syllogism in Fig. I, viz. $xm_0 \dagger ym'_0$ ⁋ xy_0.

Here our *Data* are the two Nullities, xm_0 and ym'_0 involving the Attribute m both in the *positive* and in the *negative* form: and our *Quaesitum* is the Nullity xy_0.

We begin by assuming that the aggregate xy is an *Entity:* i.e. we assume that some existing Thing has *both* the Attributes x and y.

Now the *first* Premiss tells us that x is incompatible with m. Hence the 'Thing' under consideration, which is assumed to have the Attribute x, *cannot* have the Attribute m. But it is bound to have one of the two m or m', since these constitute an *Exhaustive Division* of the whole Universe. Hence it *must* have the Attribute m'.

Similarly, from the *second* Premiss, we can prove, as our *second* result, that the 'Thing' under consideration has the Attribute m.

These two results, taken together, give us the startling assertion that this 'Thing' has *both* the Attributes, m and m', *at once;* i.e. we get xy_1 ⁋ $xym'm_1$.

Now we know that m and m' are *Contradictories:* hence this result is evidently *absurd:* so we go back to our original assumption (that the aggregate xy was an *Entity*), and we say "hence xy *cannot* be an *Entity:* that is, it is a *Nullity.*"

Now let us arrange this argument in the form of a *Tree.*

I must explain, to begin with, that all the Trees, in this system, grow *head-downwards:* the Root is at the *top*, and the Branches are *below*. If it be objected that the name "Tree" is a misnomer, my answer is that I am only following the example of all writers on *Genealogy.* A *Genealogical* 'Tree' *always* grows *downwards:* then why may not a *Logical* 'Tree' do likewise?

Well, then, I put the *Root* of my Tree at the top. It consists of the aggregate xy: and the mere writing down of these two Letters is to be understood to mean (using the regular form of a *Reductio ad Absurdum*) "The aggregate xy shall be a *Nullity:* for, if not, let it be an *Entity;* that is, let a certain existing Thing have the two Attributes, x and y."

Underneath this 'xy' I then place the Letter m' (this is part of the *Stem* of our Tree): and on its left-hand side I place the Number '1,' followed by a full-stop, so that our Tree is now ....

xy
1. m'

The meaning of this is, that the 'Thing,' which is assumed to have the two Attributes x and y, *must also* have the Attribute m': and the Number '1' refers you to the *first* Premiss as my authority for this assertion.

Next, I place the Letter m on the right-hand side of m', and the Number '2,' followed by a comma, on the *left*-hand side of the '1,' so that our Tree now is....

xy
2, 1. $m'm$

This means that the 'Thing' *must also* have the Attribute m, (i.e. that $xym'm$ is an *Entity*) and that my authority, for asserting this, is the *second* Premiss. (Observe that the two *Letters*, in the lower line, are to be read *from left to right*, but the two *Reference-Numbers from right to left.*)

Now we know that m' and m are *Contradictories:* hence it is impossible for an Aggregate, which contains them *both*, to be an *Entity:* hence it is a *Nullity.* And *this* fact I indicate by drawing a little *circle* (representing a *nought*) underneath, so that our Tree now is..................

xy
2, 1. $m'm$
∴ ○

The meaning of the circle is "The aggregate of Attributes, beginning at the Root, down to this point, is a *Nullity.*"

Next, I place, underneath the little circle, the Conclusion "∴ xy_0," so that Tree now is..................

xy
2, 1. $m'm$
○
∴ xy_0

The meaning of the last line is "We have now proved, from the assumption that xy was an *Entity*, that this aggregate, $xym'm$, must be an *Entity*. But it is evidently a *Nullity.* Which is *absurd*. Hence our assumption was *false.* Hence we have a right to say "Therefore xy is a *Nullity.*"

METHOD OF TREES was developed by Dodgson as a means of testing the validity of a conclusion derived from premises. It is strikingly similar to the "trees" frequently used by contemporary logicians. The basic idea is to assume that the premises are true but the conclusion is false (and its denial is true). If combining the denial of the conclusion with the premises leads to an absurdity, it shows that the premises do indeed prove the conclusion. The explanation here has been reset from the section describing "The Method of Trees" in Dodgson's unpublished work. It deals with a simple tree that does not require branching.

Another paradox from the unpublished papers deals with the ancient problem of the crocodile and the baby. Dodgson wrote:

"This tragical story runs as follows:—

"A Crocodile had stolen a Baby off the banks of the Nile. The Mother implored him to restore her darling. 'Well', said the Crocodile, 'if you say truly what I shall do, I will restore it: if not, I will devour it'.

" 'You will devour it!' cried the distracted Mother.

" 'Now', said the wily Crocodile, 'I *cannot* restore your Baby: for, if I do, I shall make you speak *falsely:* and I warned you that, if you spoke *falsely*, I would *devour* it'.

" 'On the contrary', said the yet wilier Mother, 'you cannot *devour* my Baby: for, if you do, you will make me speak *truly*, and you promised me that, if I spoke *truly*, you would *restore* it!' (We assume, of course, that he was a Crocodile of his word; and that his sense of honour outweighed his love of Babies.)"

Dodgson then applies his subscript logic and his "Method of Trees" to the problem:

"On this Sophism [R. H.] Lotze makes the discouraging remark, 'There is no way out of this dilemma.' I think, however, that we shall find the machinery of Symbolic Logic sufficient for its solution.

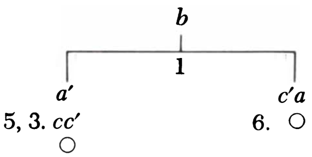
"Let Univ. be 'Cosmophases'; $a =$ the Mother speaks truly; $b =$ the Crocodile keeps his word; $c =$ the Crocodile devours the Baby."

(To negate a term or statement, Dodgson placed a prime sign immediately after it. Thus a' stands for "The Mother speaks falsely." The subscript 1 asserts the existence of the term and the subscript 0 asserts nonexistence. Thus ab_1c_0 meant "All ab are not-c," which in this particular example means: "Whenever the Mother speaks truly and the Crocodile keeps his word, then the Crocodile fails to devour [restores] the Baby." And c'_1a_0 means "All not-c are not-a," or "Whenever the Crocodile restores the Baby, then the Mother speaks falsely.")

"Then we have, as *Data*,

1. ab_1c_0	4. $a'b'_1c_0$
2. $ab'_1c'_0$	5. $c_1a'_0$
3. $a'b_1c'_0$	6. c'_1a_0

"Here we may ignore 2, 4, as being contained in 6, 5; and we see, by inspection, that b is the only Retinend.



$\therefore$ b_0: i.e. 'Whatever the Crocodile does, he *breaks* his word.'

"Thus, if he devours the Baby, he makes her speak truly, and so *breaks* his word; and if he restores it, he makes her speak falsely, and so *breaks* his word. His sense of honour being thus hopeless of satisfaction, we cannot doubt that he

BARBERSHOP PARADOX was published in the British philosophical journal *Mind* in July, 1894, but Dodgson clearly intended it for the second volume of *Symbolic Logic* along with his resolution of it. The version here is from the unpublished galley proofs.

"What, *nothing* to do?" said Uncle Jim. "Then come along with me down to Allen's. And you can just take a turn while I get myself shaved."

"All right," said Uncle Joe. "And the Cub had better come too, I suppose?"

The "Cub" was *me*, as the reader will perhaps have guessed for himself. I'm turned *fifteen*—more than three months ago; but there's no sort of use in mentioning *that* to Uncle Joe: he'd only say "Go to your cubbicle, little boy!" or "Then I suppose you can do cubbic equations?" or some equally vile pun. He asked me yesterday to give him an instance of a Proposition in *A*. And I said "All uncles make vile puns". And I don't think he liked it. However, that's neither here nor there. I was glad enough to go. I *do* love hearing those uncles of mine "chop logic," as they call it; and they're desperate hands at it, *I* can tell you!

"That is not a logical inference from my remark," said Uncle Jim.

"Never said it was," said Uncle Joe: "it's a *Reductio ad Absurdum*."

"An *Illicit Process of the Minor!*" chuckled Uncle Jim.

That's the sort of way they always go on, whenever *I'm* with them. As if there was any fun in calling me a minor!

After a bit, Uncle Jim began again, just as we came in sight of the barber's. "I only hope *Carr* will be at home," he said. "Brown's so clumsy. And Allen's hand has been shaky ever since he had that fever."

"Carr's *certain* to be in," said Uncle Joe.

"I'll bet you sixpence he *isn't!*" said I.

"Keep your bets for your betters," said Uncle Joe. "I mean"—he hurried on, seeing by the grin on my face what a slip he'd made—"I mean that I can *prove* it logically. It isn't a matter of *chance*."

"Prove it *logically!*" sneered Uncle Jim. "Fire away, then! I defy you to do it!"

"For the sake of argument," Uncle Joe began, "let us assume Carr to be *out*. And let us see what that assumption would lead to. I'm going to do this by *Reductio ad Absurdum*."

"Of course you are!" growled Uncle Jim. "Never knew any argument of *yours* that didn't end in some absurdity or other!"

"Unprovoked by your unmanly taunts," said Uncle Joe in a lofty tone, "I proceed. Carr being out, you will grant that, if Allen is *also* out, *Brown* must be at home?"

"What's the good of *his* being at home?" said Uncle Jim. "I don't want *Brown* to shave me! He's too clumsy."

"Patience is one of those inestimable qualities——" Uncle Joe was beginning; but Uncle Jim cut him off short.

"*Argue!*" he said. "Don't *moralise!*"

"Well, but *do* you grant it?" Uncle Joe persisted. "Do you grant me that, if Carr is out, it follows that if Allen is out Brown *must* be in?"

"Of course he must," said Uncle Jim; "or there'd be nobody to mind the shop."

"We see, then, that the absence of Carr brings into play a certain Hypothetical, whose *protasis* is 'Allen is out,' and whose *apodosis* is 'Brown is in'. And we see that, so long as Carr remains out, this Hypothetical remains in force?"

"Well, suppose it does. What then?" said Uncle Jim.

"You will also grant me that the truth of a Hypothetical—I mean its *validity* as a logical *sequence*—does not in the least depend on its *protasis* being actually *true*, nor even on its being *possible*. The Hypothetical 'If you were to run from here to London in five minutes you would surprise people,' remains true as a *sequence*, whether you can do it or not."

"I *ca'n't* do it," said Uncle Jim.

"We have now to consider *another* Hypothetical. What was that you told me yesterday about Allen?"

"I told you," said Uncle Jim, "that ever since he had that fever he's been so nervous about going out alone, he always takes Brown with him."

"Just so," said Uncle Joe. "Then the Hypothetical 'if Allen is out Brown is out' is *always* in force, isn't it?"

"I suppose so," said Uncle Jim. (He seemed to be getting a little nervous, himself, now.)

"Then, if Carr is out, we have *two* Hypotheticals, 'if Allen is out Brown is *in*' and 'If Allen is out Brown is *out*,' in force at once. And two *incompatible* Hypotheticals, mark you! They ca'n't *possibly* be true together!"

"*Ca'n't* they?" said Uncle Jim.

"How *can* they?" said Uncle Joe. "How *can* one and the same *protasis* prove two contradictory *apodoses?* You grant that the two *apodoses*, 'Brown is *in*,' and 'Brown is *out*,' are contradictory, I suppose?"

"Yes, I grant *that*," said Uncle Jim.

"Then I may sum up," said Uncle Joe. "If Carr is out, these two Hypotheticals are true together. And we know that they *cannot* be true together. Which is absurd. Therefore Carr *cannot* be out. There's a nice *Reductio ad Absurdum* for you!"

Uncle Jim looked thoroughly puzzled; but after a bit he plucked up courage, and began again. "I don't feel at all clear about that *incompatibility*. Why shouldn't those two Hypotheticals be true together? It seems to me that would simply prove '*Allen* is in.' Of course it's clear that the *apodoses* of those two Hypotheticals are incompatible—'Brown is in' and 'Brown is out'. But why shouldn't we put it like this? If Allen is out Brown is *out*. If Carr and Allen are *both* out, Brown is *in*. Which is absurd. Therefore Carr and Allen ca'n't be *both* of them out. But, so long as Allen is *in*, I don't see what's to hinder Carr from going *out*."

"My dear, but most illogical, brother!" said Uncle Joe. (Whenever Uncle Joe begins to "dear" you, you may make pretty sure he's got you in a cleft stick!) "Don't you see that you are wrongly dividing the *protasis* and the *apodosis* of that Hypothetical? Its *protasis* is simply 'Carr is out'; and its *apodosis* is a sort of sub-Hypothetical, 'If Allen is out, Brown is *in*'. And a most absurd apodosis it is, being hopelessly incompatible with that other Hypothetical, that we know is *always* true, 'If Allen is out, Brown is *out*'. And it's simply the assumption 'Carr is out' that has caused this absurdity. So there's only *one* possible conclusion. *Carr is in!*"

How long this argument *might* have lasted, I haven't the least idea. I believe *either* of them could argue for six hours at a stretch. But, just at this moment, we arrived at the barber's shop; and, on going inside, we found——

would act in accordance with his *second* ruling passion, his love of Babies!

"[The Reader will find it an interesting exercise to work out for himself the result which would have followed if the Mother's first statement had been 'You will *restore* the Baby.' He will find that, in that case, whatever the Crocodile does, he *keeps* his word. Hence his sense of honour is entirely satisfied, whatever he does: so that, again, his only guide is his *second* ruling passion—and the result to the Baby would, I fear, be much the same as before.]"

In his comment that data 2 and 4 are contained in data 6 and 5, Dodgson meant that the former are logically derivable from the latter. The conclusion reached by applying Dodgson's rule of inference is: "No b exist," or "There are no cases in which the Crocodile keeps his word," or in Dodgson's own terms, "Whatever the Crocodile does, he breaks his word."

The logic tree is an application of a *reductio ad absurdum* argument to the hypothetical assumption that b_0 (no b exist) is false. We postulate b_1 (b do exist). This information, together with the first datum (or premise), allows for two possibilities. The first, in the left-hand branch of the tree, is a' (the Mother speaks falsely). This result, however, together with b, leads by the third premise to c, whereas by the fifth premise it leads to c'. Since c and c' are contradictory, the first possibility leads to an absurdity (indicated by the circle). The second possibility, shown in the right-hand branch of the tree, of conjoining b_1 conjoined with the first premise produces $c'a$ (the Crocodile restores the Baby and the Mother speaks truly). This conflicts with the sixth premise and also leads to an absurdity. If the assumption that the Crocodile can sometimes keep his word always leads to absurdity, then indeed "Whatever the Crocodile does, he breaks his word."

To test our command of Dodgson's technique, let us attempt to determine what happens if the Mother says: "You will restore the Baby." Datum 5 becomes c_1a_0 (all c are not-a), and datum 6 becomes $c'_1a'_0$ (all not-c are a). That is, "If the Crocodile devours the Baby, then the Mother speaks falsely," and "If the Crocodile restores the Baby, the Mother speaks truly." Here data 1 and 3 can be disregarded because they are derivable from data 5 and 6. The relevant data are then 2, 4, 5 and 6. The conclusion is b'_0, "Whatever the Crocodile does he keeps

" A Logical Puzzle"
By Lewis Carroll

There are three Propositions, A, B, and C.
It is given that

"If A is true, B is true; ....(i)
"If C is true, then if A is true B is not true ....(ii)

NEMO and OUTIS differ about the truth of C.
NEMO says C cannot be true; OUTIS says it can.

NEMO's Argument.

Number (ii) amounts to this:
"If C is true, then (i) is not true".
But, *ex hypothesi*, (i) *is* true".
∴ C cannot be true; for the assumption of C involves an absurdity.

OUTIS's Reply.

NEMO's two assertions, "if C is true, then (i) is not true" and "the assumption of C involves an absurdity", are erroneous.

The assumption of C *alone* does *not* involve any absurdity, since the two Hypotheticals, "if A is true B is true" and "if A is true B is not true", are *compatible;* i.e. they can be true together, in which case A cannot be true.

But the assumption of C and A *together does* involve an absurdity; since the two Propositions, "B is true" and "B is not true", are *incompatible*.

Hence it follows, not that C, *taken by itself*, cannot be true, but that C and A cannot be true *together*.

NEMO's Rejoinder.

OUTIS has wrongly divided Protasis and Apodosis in (ii).

The absurdity is not the last clause of (ii), "B is not true", but *all* that follows the word "then", i.e. the Hypothetical "If A is true B is not true"; and, by (ii), it is the assumption of C only which causes this absurdity.

In fact, OUTIS has made (ii) equivalent to "If C is true [and if A is true] then if A is true B is not true". This is erroneous: the words in the brackets in the true compound Protasis are superfluous, and the remainder is the true Protasis which conditions the absurd Apodosis, as is evident from the form of (ii) originally given.

This Theorem in Hypotheticals—that the Propositions, numbered (i) and (ii), together prove that C cannot be true—may be illustrated by the following algebraical example:—

Let $ax + (a - b)y + z = 5$; .... (1)
$bx + z = 6$; .... (2)

Equation (1) may be stated as a Hypothetical, thus:—
"If ax, $(a - b)y$, and z be added together, the number '5' is obtained".

Let 'A' mean "ax, $(a - b)y$, and z are added together";
'B' " "the number '5' is obtained";
'C' " "$a = b$".

Then we have
"If A is true, B is true".
Assume that C is true; i.e. that $a = b$.
Then $(ax + (a - b)y + z)$ becomes $(bx + z)$, which, by Equation (2), must *always* = 6.
Hence
"If C is true, then if A is true B is not true".
Therefore C cannot be true;
i.e. 'a' cannot = 'b'.

OUTIS's Second Reply.

This reply will include (α) a proof that "NEMO's Argument" is self-destructive; (β) a proof that his algebraical example fails, owing to its not correctly representing the data; (γ) a proof that, when corrected, it illustrates OUTIS's contention, viz. that Hypotheticals (i) and (ii) prove, not that C, *taken by itself*, cannot be true, but that C and A cannot be true *together;* (δ) a simple proof of the *true* outcome of these two Hypotheticals.

(α)

Let us consider the Trio of Hypotheticals (which we will call (K), (L), and (M))

(K) "If X is true, Y is not true".
(L) "If X is true, Y is true".
(M) "If X is not true, Y is true".

It will not be disputed that (L) and (M), taken together, are equivalent to the Categorical (which we will call 'N') "Y is true." Hence the above Trio of Hypotheticals is equivalent to the Hypothetical and Categorical

(K) "If X is true, Y is not true".
(N) "Y is true".

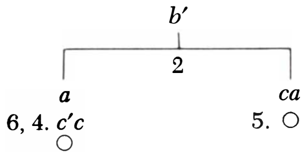
For this Trio (Or its equivalent Pair) two different interpretations might be proposed, viz.

"(K) and (L) cannot be true together. Hence, (K), (L), and (M) cannot be true together."

"(K) and (N) can be true together; that is, (K), (L), and (M) can be true together."

RESOLUTION OF BARBERSHOP PARADOX found in Dodgson's papers was his final treatment of the problem. He restates the paradox in more abstract terms and goes on to give pro and con arguments about whether or not C can be true ("Carr is out" in the original version). Nemo (Uncle Joe) presents the view held by John Cook Wilson, and Outis (Uncle Jim) the view held by Dodg-

his word." To test the argument by Dodgson's logic-tree method, assume that b'_0 is false and test whether or not this leads to absurdity. If b'_0 is false, then b'_1 is true, and this together with the second premise (datum) allows for two possibilities.



The first (the left branch of the tree) is a' (the Mother speaks falsely), but $a'b'_1$ by the fourth premise leads to c' and by the sixth premise to c. Having c' and c together is absurd. The second possibility is ca_1 (the right branch), which contradicts the fifth premise and so is absurd. Thus b'_0 must be true and whatever the crocodile does, he keeps his word.

Although the crocodile dilemma was not originated by Dodgson (it goes back at least to the Stoic logic of ancient Greece), I have never seen its full implications so precisely—and movingly—set out. Still, a lingering doubt remains: Dodgson maintained he had solved the dilemma. What dilemma if any, did he solve (as opposed to analyzing)? Indeed, precisely what is the dilemma of the crocodile? I leave that problem with you: the proof of the spoof is in the putting.

Another problem that Dodgson takes up is the famous Paradox of the Liar. He reports its "simplest form" thus: "If a man says 'I am telling a lie,' and speaks truly, he *is* telling a lie, and therefore speaks falsely: but if he speaks false ly, he is *not* telling a lie, and therefore speaks truly."

Many logicians have in recent years dismissed the Paradox of the Liar out of hand, on the ground that the paradox arises from allowing self-reference, that is, allowing sentences to refer to their own truth or falsity. Such an attitude probably arises from a too hasty reading of Alfred Tarski's 1931 paper "The Concept of Truth in Formalized Languages," where Tarski argues that no consistent language can contain the means for speaking of the meaning or the truth of its own expressions. When a language does allow self-reference, it is not surprising that it will lead to inconsistencies and paradoxes.

Dodgson himself takes up such a sug-

These interpretations are *incompatible.*

Now, when NEMO says "the assumption of C involves an absurdity", the "absurdity", to which he alludes, is the *simultaneous truth* of the two Propositions "If A is true B is true" and "If A is true B is not true".

These two Propositions are Hypotheticals of the forms (L) and (K): and, in declaring that the assumption of their simultaneous truth involves an absurdity, NEMO virtually declares that they *cannot be true together.*

Here, then, he adopts the *first* interpretation of the Trio of Hypotheticals, (K), (L), and (M).

Again, when he says "$\therefore$ C cannot be true", the premisses, from which he deduces this conclusion, are the two Propositions "If C is true, then (i) is not true. But, *ex hypothesi*, (i) *is* true."

These two Propositions are a Hypothetical and a Categorical of the forms (K) and (N): and, in deducing a conclusion from them, regarded as premisses, NEMO virtually declares that they *can be true together.*

Here, then, he adopts the *second* interpretation of the Trio of Hypotheticals, (K), (L), and (M).

Thus he has adopted, in the course of one and the same argument, *two incompatible* interpretations of this Trio.

Hence "NEMO's Argument" is self-destructive.

(β)

Let us now examine NEMO's algebraical example.

He gives us Equations (1) and (2) as *always true.*

Hence Equation (1) remains true, even when $a = b$.

Hence, his second Hypothetical is incomplete: it ought to be "If C is true, then if A is true B is (by Equation 1) true, but (by Equation 2) not true".

Hence his algebraical example fails, owing to its not correctly representing the data.

(γ)

The two Hypotheticals, when *fully* stated, run thus:—

"If A is true, B is (by Equation 1) true";

"If C is true, then if A is true B is (by Equation 1) true, but (by Equation 2) not true."

These two may be stated as *three* Hypotheticals, viz.

"If A is true, B is (by Equation 1) true";

"If C is true, then if A is true B is (by Equation 1) true;

"If C is true, then if A is true B is (by Equation 2) not true".

The second of these we may omit, as it leads to no result. The other two may be more briefly stated thus:—

"If A and (1) are true, B is true;

If C and A and (2) are true, B is not true".

And the correct conclusion is, not that C, *taken by itself*, cannot be true, but that C, A, (1), and (2) cannot all be true *together.*

But A is *always possible*; so that we may, if we like, assume it as *always true*, and not mention it.

The two Hypotheticals may now be written thus:—

"If (1) is true, B is true;

If C and (2) are true; B is not true".

Therefore C and (1) and (2) cannot all be true *together*, though any *two* of them may be true *by themselves.*

Thus, if C and (1) are true, then (2) cannot be true: that is, if $a = b$ (so that Equation 1 becomes "$bx + z = 5$"), and if Equation 1 is true, then it cannot be true that $bx + z = 6$.

Secondly, if C and (2) are true, then (1) cannot be true: that is, if $a = b$, and if $bx + z = 6$, then it cannot be true that $ax + (a - b)y + z = 5$.

Thirdly, if (1) and (2) are true, then C cannot be true: that is, if *both* the given Equations are true, then a cannot $= b$.

This algebraical example might easily mislead an unwary reader, from the fact that its Conclusion, "C cannot be true," is (on the assumption that Equations 1 and 2 are always true) a true one. The fallacy lies in prefixing the word "Therefore," and thereby asserting that this Conclusion *follows from the two Hypotheticals.* This is *not* the case: the *real* reason, why C cannot be true, is that *it is incompatible with Equations 1 and 2* (by subtraction we get $(a - b)(x + y) = -1$, whence it follows that $(a - b)$ cannot $= 0$; i.e. that a cannot $= b$); the two Hypotheticals, by themselves, do *not* prove it.

(δ)

The *true* outcome, of the original Hypotheticals numbered (i) and (ii), may be very simply exhibited as follows:—

Let '*t*' stand for "true", and '*f*' for "false".

There are 8 conceivable combinations of A, B, and C, with regard to truth and falsity: these are as follows:—

	1.	2.	3.	4.	5.	6.	7.	8.
A.	*t*	*t*	*t*	*t*	*f*	*f*	*f*	*f*
B.	*t*	*t*	*f*	*f*	*t*	*t*	*f*	*f*
C.	*t*	*f*	*t*	*f*	*t*	*f*	*t*	*f*

Of these, Nos. 3 and 4 are forbidden by (i), and No. 1 is forbidden by (ii).

The other 5 combinations are *possible*; and *two* of them, viz. Nos. 5 and 7, contain the condition "C is true", which NEMO believes to be *impossible.*

[*September*, 1894]

son. Throughout the dispute Dodgson presses the point that a conditional is false only when the antecedent is true and the conclusion false. The truth table near the end is used to show that there are possible situations in which C is true, contrary to Nemo's arguments. Even so, this paradox with its conflicting hypotheticals raised issues that are still the subject of controversy among logicians today.

gestion, considers it seriously and then rejects it, all in the space of a few lines. He writes:

"The best way out of the difficulty [the Paradox of the Liar] seems to be to raise the question whether the Proposition 'I am telling a lie' can reasonably be supposed to refer to *itself* as its own subject-matter."

He concluded that "I am telling a lie" may indeed not be permitted to refer to itself, "since doing so would lead to an absurdity." He did not stop there. He went on to stress that self-reference in and by itself is not objectionable, and he pointed out that a man's statement "I am telling the truth" does not lead to an absurdity when it refers to itself.

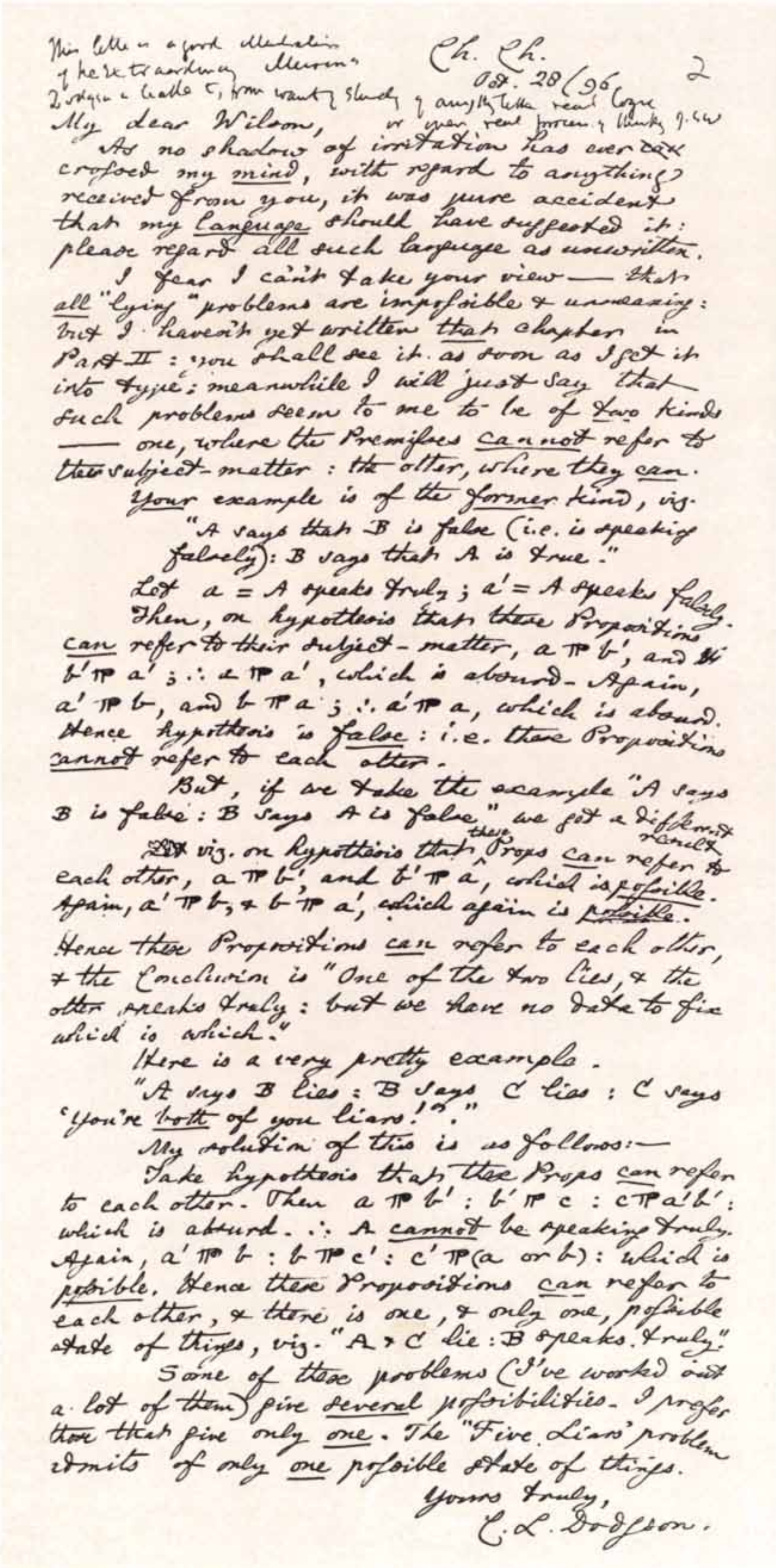
The point, which Dodgson saw clearly, is that some self-referential statements cause no difficulty, whereas others engender paradox. Consider:

The sentence in this box is true.

The sentence in this box is false.

Alan Ross Anderson of the University of Pittsburgh recently summed up the situation that would result if some particular types of self-reference were not allowed. He wrote: "We would lose virtually all of the most interesting fields in contemporary studies in the philosophical foundations of mathematics. The fundamental theorems of set theory and of recursion theory would disappear, and mathematicians and logicians the world over would be out of business."

How will publication of the second part of Charles Lutwidge Dodgson's *Symbolic Logic* affect contemporary logic? Some logicians will take great interest in tracing out the evolution of Dodgson's decision procedure and his validity tests; others will enjoy his witty analysis of paradoxes and puzzles. The greatest impact, I believe, will be made by the 80 new exercises in the book. For more than half a century logicians have been stealing for their own textbooks and classroom notes the zany problems Dodgson set out in the first volume of *Symbolic Logic*. They have on the whole had an easy time of it, since Dodgson provided answers to the exercises in that book. Answers for a few of the new exercises have been found, mainly in Dodgson's correspondence with his sisters and with Wilson. For most of the problems no solutions are given in the surviving text. If Lewis Carroll could watch us puzzle over them, I suspect he would *chortle* with delight.



This letter is a good illustration of the extraordinary illusions Dodgson is liable to, from want of study of anything like real logic or even real process of thinking. J. C. W.

Ch. Ch. 2

Oct. 28/96

My dear Wilson,

As no shadow of irritation has ever crossed my mind, with regard to anything received from you, it was pure accident that my language should have suggested it: please regard all such language as unwritten.

I fear I can't take your view — that all "lying" problems are impossible & unmeaning: but I haven't yet written that chapter in Part II: you shall see it as soon as I get it into type; meanwhile I will just say that such problems seem to me to be of two kinds — one, where the Premisses cannot refer to their subject-matter: the other, where they can.

Your example is of the former kind, viz.

"A says that B is false (i.e. is speaking falsely): B says that A is true."

Let a = A speaks truly; a' = A speaks falsely.

Then, on hypothesis that these Propositions can refer to their subject-matter, a IP b', and b' IP a'; ∴ a IP a', which is absurd. Again, a' IP b, and b IP a; ∴ a' IP a, which is absurd. Hence hypothesis is false: i.e. these Propositions cannot refer to each other.

But, if we take the example "A says B is false: B says A is false" we get a different result viz. on hypothesis that these Props can refer to each other, a IP b', and b' IP a, which is possible. Again, a' IP b, & b IP a', which again is possible. Hence these Propositions can refer to each other, & the Conclusion is "One of the two lies, & the other speaks truly: but we have no data to fix which is which."

Here is a very pretty example.

"A says B lies: B says C lies: C says 'you're both of you liars!'."

My solution of this is as follows:—

Take hypothesis that these Props can refer to each other. Then a IP b' : b' IP c : c IP a'b' : which is absurd. ∴ A cannot be speaking truly. Again, a' IP b : b IP c' : c' IP (a or b): which is possible. Hence these Propositions can refer to each other, & there is one, & only one, possible state of things, viz. "A & C lie: B speaks truly."

Some of these problems (I've worked out a lot of them) give several possibilities. I prefer those that give only one. The "Five Liars" problem admits of only one possible state of things.

Yours truly,

C. L. Dodgson.

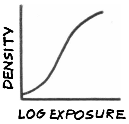
LIAR PARADOX was the subject of correspondence dated October 28, 1896, between Dodgson and Wilson. At top of letter Wilson has scrawled: "This letter is a good illustration of the extraordinary illusions Dodgson is liable to, from want of study of anything like real logic or even real process of thinking. J. C. W." *Ch. Ch.* stands for Christ Church.

We want to be useful ...and even interesting

The difference between electrons and light

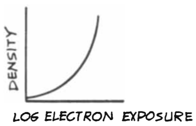
Many a working biologist knows more than we ever knew about how to make a good transmission electron micrograph.

An elementary photography course teaches



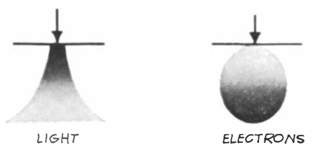
about the classic density-log exposure curve, where the slope of the linear portion is a measure of contrast and is controlled by the degree of development.

Later one learns that for exposure by electrons instead of light, the working curve changes to this shape. With this shape, for more contrast one

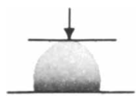


simply gives more exposure, with due regard for the danger of frying the biostructures under study. With more exposure comes less graininess. But this is not the graininess they were talking about in the elementary photography course. It is quantum mottle. To create a given density takes a lot fewer 50-kV electrons than photons of sunshine. Paint the picture with more electrons and you see less random fluctuation.

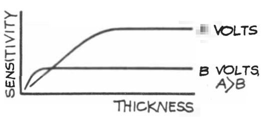
Another difference:



The electrons have to slow down from inelastic collisions before they can have much photographic effect. If the emulsion is too thin for the kilovoltage, it will



act photographically slow because they are zipping right on through. Hence this principle:



KODAK Electron Image Plates differ from the old Projector Slide Plates by a thicker emulsion. A beneficial side effect yields more uniform sensitivity from center to edge. No apologies, however, are offered for the old Projector Slide Plates, even though they were indistinguishable from the still older Lantern Slide Plates, renamed some years after it became customary to refer to a magic lantern as a projector. They captured some important information about the machinery of life during the earlier, lower-kilovoltage days of EM.

The information package "Kodak Products for Electron Microscopy," from Dept. 412-L, Kodak, Rochester, N.Y. 14650, may prove helpful with operating suggestions. It tells about film, too.

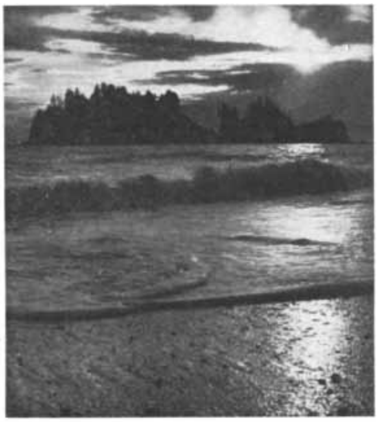
512 colors

It seems possible to detect and identify substances at a distance by a highly specialized kind of color photography, amounting to abridged spectrophotometry. The film would have separate, relatively narrow response bands controlling the three (or more?) dye layers in the end product. For other purposes it would hardly be worth its price. No orders accepted. Not yet anyway. If ever.

"Very slightly different colors, the distinction of which is possible only when two of them occupy large, sharply contiguous areas, can convey information the importance of which may far transcend the information-theory measure of its quantity."

So writes David L. MacAdam, a Kodak man whose surname was long ago draped by the world's psychophysicists around the unit of "least distinguishable difference" in color perception. He is of the opinion that "for the same visual conditions under which there can be sixteen distinctly different grays, 512 equally distinct colors (including those grays) can be produced in a color photograph."

The problem of accessibility



A hundred nations now maintain national parks. The idea originated in, of all places, the U.S.A. To commemorate the centennial of the first national park, we have made six films and presented them to the National Park Service, which is delighted to have them shown on TV. Perhaps you have caught one or soon will.

Our regular business is to make film. It is our customers who make films. In the case of these six films, exception is justified on the grounds of stimulating travel, which in turn and by tradition stimulates picture-taking. But there lurks a problem. Lack of park patronage is not the problem.

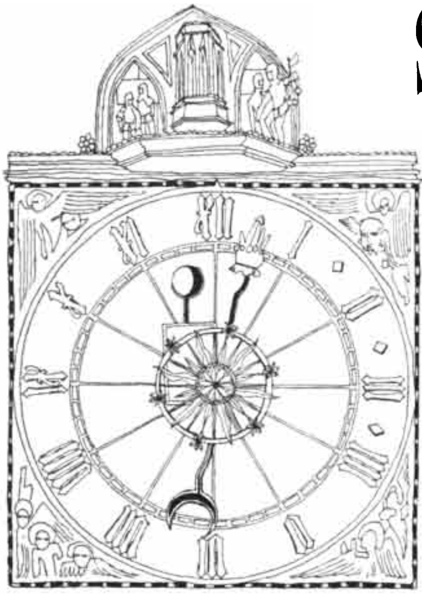
Use of the parks concentrates in July and August, the 1/6 of the year when schoolkids get vacation to help with the farming, even if the family has not actually farmed for four or five generations. Very little of our beautiful footage was shot during those months. We were allowed in during the other 5/6 of the year, and so are you. If your own obligations permit, and if you are willing to miss the opportunity to photograph masses of other people's kids licking ice cream cones, you may find other acceptable camera fodder.

Did you know that the Yellowstone elk spend all summer in the generally inaccessible high country and that wild geese stay all winter near the famous hydrothermals that make that old park the world's most generally accessible demonstration of super-speed geology?

Kodak

Imagery for purposes new and old

SCIENCE AND



Quantity Control

The outcome of the recently concluded first phase of the strategic-arms-limitation talks (SALT I) between the U.S. and the U.S.S.R. has been widely heralded as a "historic first step" and a "major achievement" on the road toward curbing the nuclear arms race. For the first time in the postwar period the two superpowers have agreed to accept quantitative ceilings on their respective arsenals of offensive and defensive missiles. Moreover, as Richard L. Garwin remarks in his article on the interaction of antisubmarine warfare and national security in this issue of SCIENTIFIC AMERICAN, the pair of accords signed in Moscow on May 26 by President Nixon and Secretary Brezhnev constitute the first formal recognition "that the two countries accept the military and political value of mutual deterrence as a means of forestalling an all-out nuclear war." (Curiously the actual texts of the two documents in question—a treaty limiting ABM, or anti-ballistic-missile, systems and an interim agreement freezing offensive missiles at roughly their present levels—make no mention of the deterrence doctrine.)

In spite of the obvious gains registered by the new pacts, concern has been expressed in various quarters since the formal signing ceremony over the purely quantitative nature of the limitations imposed by the agreements. It has been suggested that the arms race will simply be diverted from a quantitative competition to a qualitative one, yielding little actual economic savings for either country. Recent public statements by spokesmen for the Nixon Administration, notably Secretary of Defense Laird, have tended to exacerbate these fears by stressing the demands of the Joint Chiefs of Staff for new missile-submarine and strategic-bomber programs as a condition of their support for the SALT I agreements. Thus it appears likely that the arms-control debate in Congress will focus not so much on what the SALT agreements prohibit as on what they permit, and in particular on what new weapons programs, if any, need to be undertaken in the new strategic situation created by the accords. By extension, the debate will also be addressing the question of what the agenda of the next round of strategic-arms-limitation talks, SALT II, should or should not include.

Of the two SALT I agreements, the ABM treaty is comparatively straightforward and potentially less controversial; it is expected to meet negligible opposition to ratification in the Senate. Basically the treaty limits each party to a total of 200 operational defensive-missile launchers deployed at two 100-missile sites, one centered on the national capital and the other in the vicinity of an existing field of offensive-missile launchers. The treaty also limits the number, size and location of the radar systems associated with the two permitted ABM systems. In addition the treaty includes a number of other provisions aimed in general at eliminating the possibility that either party would attempt to acquire a nationwide ABM defense, the basic unstated premise being that such an attempt would threaten the viability of the mutual-deterrence principle and hence increase the risk of nuclear war.

Thus the ABM treaty permits (but by no means compels) the U.S., which already has an ABM complex nearing completion at its Minuteman intercontinental-ballistic-missile (ICBM) base at Grand Forks Air Force Base in North Dakota, to add a second, 100-missile ABM defense around Washington. The U.S.S.R., on the other hand, has already installed a partial ABM defense around Moscow and is permitted by the terms of the treaty to add a second, 100-missile ABM system to protect one of its remote ICBM fields. The treaty requires the U.S. to destroy or dismantle "within the shortest possible agreed period of time" a second ABM complex currently under construction at Malmstrom Air Force Base in Montana and prevents the establishment of the remainder of the 12 sites originally planned (but never authorized) for the Safeguard ABM system.

The interim agreement restricting each party's land-based and sea-based offensive missiles to those currently deployed or under construction consists of a complex set of numerical formulas aimed at maintaining what both sides evidently regard as the present state of parity existing between strategic forces of somewhat different composition. The agreement places no limits on qualitative improvements in offensive missiles, no ceilings on the number of warheads carried by such missiles and no controls at all on the number or types of strategic bombers permitted on either side. Moreover, during the five-year term of the interim agreement both sides are permitted to phase out certain of their older (pre-1964) ICBM's and to replace them with additional missile submarines. The U.S.S.R. would in addition be permitted to substitute modern nuclear-powered missile submarines for some of their obsolete diesel-powered missile submarines, although any new submarine missiles would have to be counted within its overall missile total.

One potentially obstructive point worked out in Moscow by President Nixon and Secretary Brezhnev was the size of the Russian missile-submarine fleet. The U.S. negotiators at the SALT sessions in Helsinki and Vienna had been proceeding on the assumption that the Russians currently have a total of 42 nuclear-powered Y-class missile submarines deployed or under construction, armed with a total of 672 sea-launched ballistic missiles (SLBM's). The Russians, however, claimed to have a total of 48 such submarines with 768 SLBM's. (The disparity in the estimates of the two sides was reportedly the result of differing definitions of what constituted a submarine under construction.) In the end the two leaders agreed to compromise on a total of 45 Russian nuclear submarines with 710 missiles. The U.S. Polaris-Poseidon fleet currently consists of 41 submarines with 656 missiles.

The two sides also accepted that at present the U.S.S.R. has 1,618 land-based ICBM's (including 313 SS-9's or other large launchers) compared with 1,054 for the U.S. (1,000 Minutemen and 54 Titans). Thus in terms of total numbers of offensive missiles the U.S.S.R. currently leads the U.S. by 2,358 to 1,710 (counting the 30 missiles currently deployed in obsolete diesel-powered submarines). If the Russians were to exercise all their options of replacing obsolete missiles and submarines, they could achieve a total of 950 SLBM's on 62 submarines with a reduction to 1,408 ICBM's, for an identical total of 2,358 missiles. Taking advantage of comparable options, the U.S. could eventually reach a total of 710 SLBM's on 44 submarines with a reduction to 1,000 ICBM's (by phasing out the Titans), thereby maintaining the total of 1,710 missiles. In any event the U.S.S.R. would still have a numerical edge of more than 600 missiles.

Offsetting the Russians' quantitative advantage in offensive missiles, however, are a number of significant U.S. advantages. For one thing, the U.S. currently has 460 strategic bombers, compared with 140 for the U.S.S.R. More important, the substantial technological head start of the U.S. in the area of multiple independently targeted reentry vehicles (MIRV's) is expected to result by July 1 in a lead in deliverable strategic nuclear weapons of 5,700 for the U.S. to 2,500 for the U.S.S.R. Finally, none of these totals includes any of the nuclear weapons available to U.S. "tactical" air units based in Europe within range of the U.S.S.R.

In short, the seeming numerical superiority of the Russian nuclear arsenal is more than matched by the technological superiority and forward-based deployment of the U.S. arsenal. The desired effect of the SALT I agreements is to institutionalize the stable mutual deterrence that is believed to reside in this state of strategic parity. The essence of what both sides appear to have accepted in signing the SALT I accords is that regardless of the internal pressures that may exist in each country for continuing the arms race either quantitatively or qualitatively, the only possible end result, insofar as each country's national security is concerned, would be merely to alter the degree of retaliatory overkill.

The New Land Camera

Edwin H. Land, founder and president of the Polaroid Corporation, gave the first extensive account of his firm's newest system of photography at a recent meeting of the Society of Photographic Scientists and Engineers in San Francisco. Under the title "Absolute One-Step Photography," Land demonstrated and described the new camera and color film that Polaroid expects to place on the market later this year. The new system, Land said, achieves a goal he had set himself more than 20 years ago: construction of a compact, automatic camera capable of making a succession of color photographs virtually as fast as the photographer can push the button.

In the present Polaroid system the color image is recorded on a negative base and is transferred to a separate receiving sheet during a one-minute development. The photographer presses the receiving sheet against the exposed negative, thereby breaking a "pod" of chemicals and initiating the development, by pulling the two-component film between a pair of steel rollers inside the camera. At the end of a minute he peels the receiving sheet, now the finished print, from the negative base and throws the base away. The speed with which a photographer can take successive pictures is limited not only by his dexterity in pulling sheets of film out of the camera but also by his ability to keep track of the developing time once the film has been removed.

The new Polaroid camera weighs 26 ounces and when folded for carrying is roughly the size of a 400-page paperback book. It incorporates a four-element lens designed by James G. Baker of Harvard University, a leading designer of advanced optical systems. The new Polaroid lens can be focused from 10 inches to infinity by moving the front element less than a quarter of an inch. After passing through the lens light travels to a mirror that was positioned at a 45-degree angle when the camera was opened. The light is reflected onto a plastic Fresnel grating that lies flat above the film but flips out of the way when the shutter is released. Light reflected from the Fresnel grating continues through a folded pathway, finally reaching the eye for focusing. The folded optical path achieves the same result as the pentaprism used in conventional through-the-lens viewing systems but weighs much less.

The entire operation of the camera is controlled by an integrated circuit containing the equivalent of 260 to 300 transistors. Among other things, the circuit measures the amount of light entering the lens during an exposure and determines when the shutter is to be closed. It then activates a tiny motor that pushes the exposed sheet of film out of the camera through a pair of rollers. If a new five-unit module of flashbulbs is being used, the electronic circuit switches from bulb to bulb in sequence, bypassing any bulb that might be defective. Power for the electronic system and camera motor is provided by a flat battery of novel design incorporated in each film pack, which means that the photographer never has to worry about or replace batteries in the camera itself. If desired, the entire 10 sheets of film in a pack can be exposed in less than 15 seconds.

As in the present film, development of the new film is initiated when it passes between rollers that break a pod of chemicals and spread the carefully measured contents evenly between the negative and the receiving sheet. Unlike the old film, the new film is an integral structure: the receiving sheet is made of a transparent plastic that remains attached to the negative base sheet. Thus there is nothing to discard. The film as it emerges from the camera is not only perfectly dry but also resistant to moisture and scratching.

In the old design the exposed negative is shielded from light by the receiving sheet when the two are pressed together and removed from the camera. In the new design light can pass freely through the receiving sheet, since it is transparent, and would instantly fog the negative if it were not for an opaque barrier that is incorporated in the developer and spread between the two sheets. The barrier consists of finely divided particles of titanium dioxide, capable of shielding the negative from a billion times more light than was required to expose it in the first place.

When the exposed film emerges from the camera, it has an overall greenish cast; no image is immediately visible. Development has started, however, and color dyes released from the negative are migrating through the opaque barrier and depositing themselves on the transparent receiving sheet. Within about a minute the image is distinct enough so that the photographer is assured he has the picture he wants. The colors continue to become more intense for another four or five minutes, reaching maximum saturation asymptotically. (The time may be shortened before the film reaches the market.)

The new color film, which has an ASA rating of 100, incorporates a new family of dyes and yields images significantly more brilliant than those produced by the present Polaroid color film. In addition, the titanium dioxide screen between the negative and the receiving sheet reflects incident light, which tends to make the pictures look as if they were illuminated from behind. In the nature of the process the new film produces an image with no visible "grain." In answer to a question, Land estimated the resolution of the new pictures to be around 40 lines per millimeter, considerably higher than one could hope to achieve in a print of the same size (3¼ by 3¼ inches) enlarged from a typical 35-mm. color transparency.

The Big Beam

The giant proton synchrotron at the National Accelerator Laboratory in Illinois has reached its design energy of 200 GeV (billion electron volts) four months ahead of schedule and is now being prepared for its initial experimental programs. The laboratory's accomplishments drew from the *CERN Courier,* a publication of the European Organization for Nuclear Research, the editorial comment that "it is a great achievement to reach the highest energy in the world from a particle accelerator, and NAL have done it in a style which is all their own." The laboratory is operated for the U.S. Atomic Energy Commission by the Universities Research Association, a consortium of 51 universities in the U.S. and one in Canada.

The Illinois machine achieved its first beam in July, 1971, but encountered difficulty with moisture in the magnets that guide the beam. By this spring the beam had achieved its design energy, although the intensity of the beam was held to 10^{11} protons per pulse instead of the design intensity of 4×10^{13}. Among several new design concepts that appear to have succeeded is a scheme whereby the energy of the beam may be extended to 500 GeV. Officials of the laboratory are confident that the beam will reach 400 GeV, although they do not venture to predict when; about 500 GeV they are more cautious. In any event, their expectation is that before long they will be able to provide beams having enough energy, intensity and reliability for significant experiments at the frontier of particle physics.

Abortion under the Law

Now that roughly a third of the states in the U.S. have eased their laws against abortion, what has been the experience with the procedure in clinical terms? A survey of 72,988 women in the U.S. who had secured legal abortions shows that whereas 10 percent developed complications of some kind (mainly bleeding or fever), only 1 percent suffered major complications. Two relatively new abortion procedures, suction and the instillation of saline solution, accounted for 93 percent of all abortions. The traditional procedure of dilatation and curettage ("D & C") accounted for 4.5 percent and hysterotomy and hysterectomy together for 2.4 percent.

The Joint Program for the Study of Abortion was conducted by Christopher Tietze and Sarah Lewit of the Population Council. Their survey covers abortions performed in 60 teaching hospitals and six clinics in 12 states and the District of Columbia from the middle of 1970 to the middle of 1971. The abortions reported were about one-seventh of all legal abortions in the U.S. during that period. Almost half of the hospitals and five of the clinics were in the state of New York. Sixty-one percent of the abortions were performed in New York, 12 percent in California and the remaining 27 percent elsewhere.

Young, single white women, pregnant for the first time, made up the largest group seeking legal abortion. They did so usually as private patients at a clinic outside their area of residence. Black women were more likely to seek abortions at a local institution. In the middle of 1970, 72 percent of the women seeking abortion were white, 22 percent were black and 6 percent belonged to other ethnic groups (mainly Puerto Rican and Oriental). By the middle of 1971 the proportions had changed: 63 percent were white, 30 percent were black and 7 percent belonged to other ethnic groups.

Women under 18 years of age tended to have late abortions; older women were more likely to have the abortion in the first trimester (12 weeks) of pregnancy. Most abortions in the first trimester were performed by suction; most at 17 weeks or later were accomplished by the saline treatment. Complication rates for abortions in the first trimester were far lower than for abortions in the second trimester.

Only six deaths were reported, corresponding to a mortality rate of 8.2 per 100,000 abortions. (The U.S. maternal death rate in childbirth is about 20 per 100,000 live births.) Three of the women died following abortion by hysterotomy, two died after the saline treatment and one after the suction treatment. The last case involved an 18-year-old who committed suicide because of guilt feelings about having "killed her baby" before she could be told that she had not been pregnant.

Complications following the suction treatment declined more than 50 percent in the one-year study period. Tietze and Lewit remark: "These gratifying declines in complication rates can undoubtedly be attributed to the greater experience and skill of the physicians and other hospital personnel dealing with the abortion patient and improved equipment in many hospitals."

Turbulent Earth

How much is the ground displaced by an earthquake of moderate severity? Measurements by investigators from the engineering geology branch of the Metropolitan Water District of Southern California show that the San Fernando earthquake near Los Angeles on February 9, 1971, produced displacements of from six to eight feet both vertically and horizontally. The earthquake had an intensity of 6.6 on the Richter scale but was in a heavily urbanized area and produced extensive damage to buildings, roads and railroad tracks. A report of the investigation, which was undertaken because the Metropolitan Water District is building or planning to build water tunnels through parts of the earthquake zone, is published in the *Geological Society of America Bulletin* by four of the agency's investigators: R. J. Proctor, R. Crook, Jr., M. H. McKeown and R. L. Moresco.

The area that experienced surface ruptures extends 12 miles (19 kilometers) from east to west and four miles (six kilometers) from north to south. The largest single vertical displacement observed in this area by the water-district group was five feet (1.5 meters), but the maximum cumulative uplift resulting from a number of smaller displacements was 7.5 feet (2.3 meters). The maximum horizontal displacement was 6.2 feet (1.9 meters). Most of the surface displacements were associated with geological faults that were known previously to such specialized agencies as the Metropolitan Water District but had not received wide public notice. Proctor and his colleagues point out that only a small part of the Santa Susana–Sierra Madre fault system was involved in the earthquake. Since all of the system is geologically active, they write, "efforts must be made to study the entire fault system in detail and to coordinate the investigation with future urban development."

THE CHEMICAL ELEMENTS OF LIFE

Until recently it was believed that living matter incorporated 20 of the natural elements. Now it has been shown that a role is played by four others: fluorine, silicon, tin and vanadium

by Earl Frieden

How many of the 90 naturally occurring elements are essential to life? After more than a century of increasingly refined investigation, the question still cannot be answered with certainty. Only a year or so ago the best answer would have been 20. Since then four more elements have been shown to be essential for the growth of young animals: fluorine, silicon, tin and vanadium. Nickel may soon be added to the list. In many cases the exact role played by these and other trace elements remains unknown or unclear. These gaps in knowledge could be critical during a period when the biosphere is being increasingly contaminated by synthetic chemicals and subjected to a potentially harmful redistribution of salts and metal ions. In addition, new and exotic chemical forms of metals (such as methyl mercury) are being discovered, and a complex series of competitive and synergistic relations among mineral salts has been encountered. We are led to the realization that we are ignorant of many basic facts about how our chemical milieu affects our biological fate.

Biologists and chemists have long been fascinated by the way evolution has selected certain elements as the building blocks of living organisms and has ignored others. The composition of the earth and its atmosphere obviously sets a limit on what elements are available. The earth itself is hardly a chip off the universe. The solar system, like the universe, seems to be 99 percent hydrogen and helium. In the earth's crust helium is essentially nonexistent (except in a few rare deposits) and hydrogen atoms constitute only about .22 percent of the total. Eight elements provide more than 98 percent of the atoms in the earth's crust: oxygen (47 percent), silicon (28 percent), aluminum (7.9 percent), iron (4.5 percent), calcium (3.5 percent), sodium (2.5 percent), potassium (2.5 percent) and magnesium (2.2 percent). Of these eight elements only five are among the 11 that account for more than 99.9 percent of the atoms in the human body. Not surprisingly nine of the 11 are also the nine most abundant elements in seawater [*see illustration on page 54*].

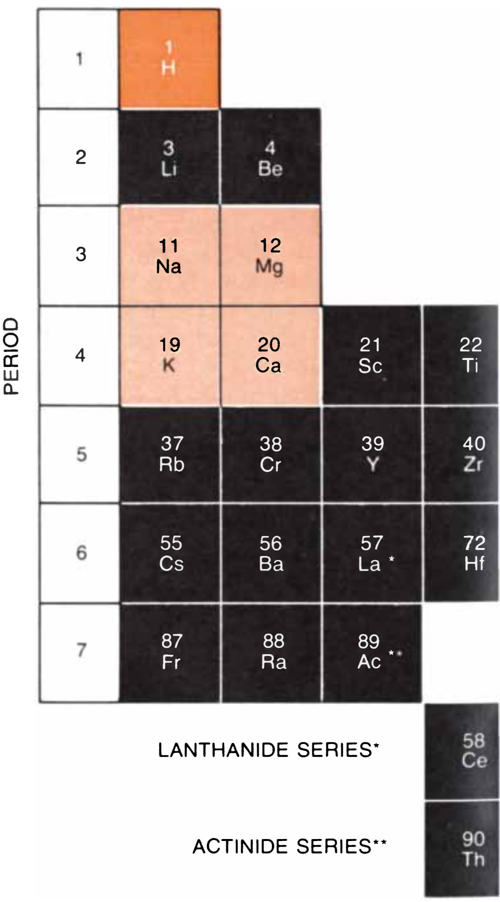
Two elements, hydrogen and oxygen, account for 88.5 percent of the atoms in the human body; hydrogen supplies 63 percent of the total and oxygen 25.5 percent. Carbon accounts for another 9.5 percent and nitrogen 1.4 percent. The remaining 20 elements now thought to be essential for mammalian life account for less than .7 percent of the body's atoms.

The Background of Selection

Three characteristics of the biosphere or of the elements themselves appear to have played a major part in directing the chemistry of living forms. First and foremost there is the ubiquity of water, the solvent base of all life on the earth. Water is a unique compound; its stability and boiling point are both unusually high for a molecule of its simple composition. Many of the other compounds essential for life derive their usefulness from their response to water: whether they are soluble or insoluble, whether or not (if they are soluble) they carry an electric charge in solution and, not least, what effect they have on the viscosity of water.

The second directing force involves the chemical properties of carbon, which evolution selected over silicon as the central building block for constructing giant molecules. Silicon is 146 times more plentiful than carbon in the earth's crust and exhibits many of the same properties. Silicon is directly below carbon in the periodic table of the elements; like carbon, it has the capacity to gain four electrons and form four covalent bonds.

The crucial difference that led to the preference for carbon compounds over silicon compounds seems traceable to two chemical features: the unusual sta-

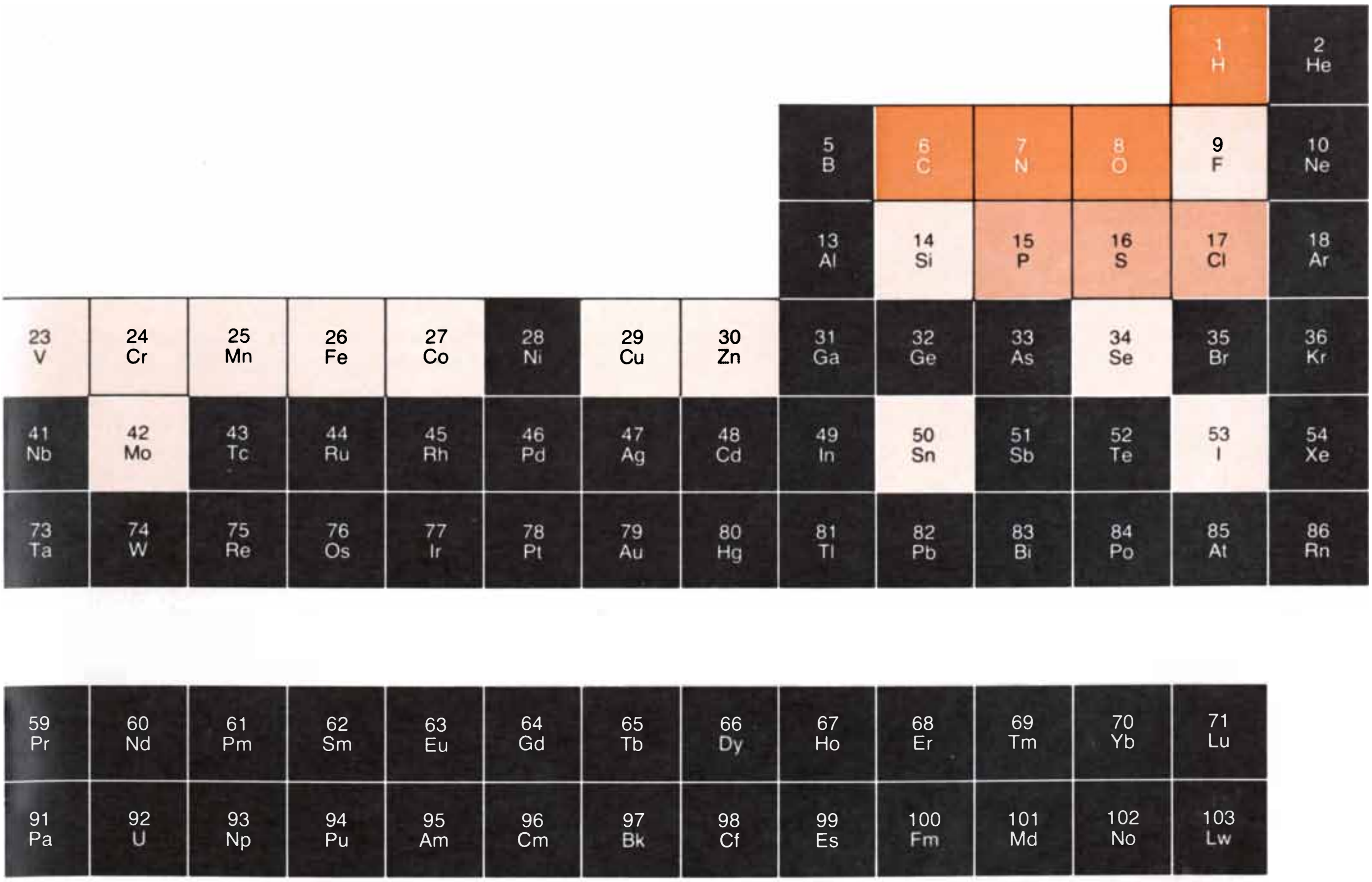


ESSENTIAL LIFE ELEMENTS, 24 by the latest count, are clustered in the upper half of the periodic table. The elements are arbility of carbon dioxide, which is readily soluble in water and always monomeric (it remains a single molecule), and the almost unique ability of carbon to form long chains and stable rings with five or six members. This versatility of the carbon atom is responsible for the millions of organic compounds found on the earth.

Silicon, in contrast, is insoluble in water and forms only relatively short chains with itself. It can enter into longer chains, however, by forming alternating bonds with oxygen, creating the compounds known as silicones (–Si–O–Si–O–Si–). Carbon-to-carbon bonds are more stable than silicon-to-silicon bonds, but not so stable as to be virtually immutable, as the silicon-oxygen polymers are. Nevertheless, silicon has recently been shown to be essential in a way as yet unknown for normal bone development and full growth in chicks.

The third force influencing the evolutionary selection of the elements essential for life is related to an atom's size and charge density. Obviously the heavy synthetic elements from neptunium (atomic number 93) to lawrencium (No. 103), along with two lighter synthetic elements, technetium (No. 43) and promethium (No. 61), were never available in nature. (The atomic number expresses the number of protons in the nucleus of an atom or the number of electrons around the nucleus.) The eight heavy elements in another group (Nos. 84 and 85 and Nos. 87 through 92) are too radioactive to be useful in living structures. Six more elements are inert gases with virtually no useful chemical reactivities: helium, neon, argon, krypton, xenon and radon. On various plausible grounds one can exclude another 24 elements, or a total of 38 natural elements, as being clearly unsatisfactory for incorporation in living organisms because of their relative unavailability (particularly the elements in the lanthanide and actinide series) or their high toxicity (for example mercury and lead). This leaves 52 of the 90 natural elements as being potentially useful.

Only three of the 24 elements known to be essential for animal life have an atomic number above 34. All three are needed only in trace amounts: molybdenum (No. 42), tin (No. 50) and iodine (No. 53). The four most abundant atoms in living organisms—hydrogen, carbon, oxygen and nitrogen—have atomic numbers of 1, 6, 7 and 8. Their preponderance seems attributable to their being the smallest and lightest elements that can achieve stable electronic configurations by adding one to four electrons. The ability to add electrons by sharing them with other atoms is the first step in forming chemical bonds leading to stable molecules. The seven next most abundant elements in living organisms all have atomic numbers below 21. In the order of their abundance in mammals they are calcium (No. 20), phosphorus (No. 15), potassium (No. 19), sulfur (No. 16), sodium (No. 11), magnesium (No. 12) and chlorine (No. 17). The remaining 10 elements known to be present in either plants or animals are needed only in traces. With the exception of fluorine (No. 9) and silicon (No. 14), the remaining eight occupy positions between No. 23 and No. 34 in the periodic table [*see illustration below*]. It is inter-



ranged according to their atomic number, which is equivalent to the number of protons in the atom's nucleus. The four most abundant elements that are found in living organisms (hydrogen, oxygen, carbon and nitrogen) are indicated by dark color. The seven next most common elements are in lighter color. The 13 elements that are shown in lightest color are needed only in traces.

esting that this interval embraces three elements for which evolution has evidently found no role: gallium, germanium and arsenic. None of the metals with properties similar to those of gallium (such as aluminum and indium) has proved to be useful to living organisms. On the other hand, since silicon and tin, two elements with chemical activities similar to those of germanium, have just joined the list of essential elements, it seems possible that germanium too, in spite of its rarity, will turn out to have an essential role. Arsenic, of course, is a well-known poison.

Functions of Essential Elements

Some useful generalizations can be made about the role of the various elements. Six elements—carbon, nitrogen, hydrogen, oxygen, phosphorus and sulfur—make up the molecular building blocks of living matter: amino acids, sugars, fatty acids, purines, pyrimidines and nucleotides. These molecules not only have independent biochemical roles but also are the respective constituents of the following large molecules: proteins, glycogen, starch, lipids and nucleic acids. Several of the 20 amino acids contain sulfur in addition to carbon, hydrogen and oxygen. Phosphorus plays an important role in the nucleotides such as adenosine triphosphate (ATP), which is central to the energetics of the cell. ATP includes components that are also one of the four nucleotides needed to form the double helix of deoxyribonucleic acid (DNA), which incorporates the genetic blueprint of all plants and animals. Both sulfur and phosphorus are present in many of the small accessory molecules called coenzymes. In bony animals phosphorus and calcium help to create strong supporting structures.

The electrochemical properties of living matter depend critically on elements or combinations of elements that either gain or lose electrons when they are dissolved in water, thus forming ions. The principal cations (electron-deficient, or positively charged, ions) are provided by four metals: sodium, potassium, calcium and magnesium. The principal anions (ions with a negative charge because they have surplus electrons) are provided by the chloride ion and by sulfur and phosphorus in the form of sulfate ions and phosphate ions. These seven ions maintain the electrical neutrality of body fluids and cells and also play a part in maintaining the proper liquid volume of the blood and other fluid systems. Whereas the cell membrane serves as a physical barrier to the exchange of large molecules, it allows small molecules to pass freely. The electrochemical functions of the anions and cations serve to maintain the appropriate relation of osmotic pressure and charge distribution on the two sides of the cell membrane.

One of the striking features of the ion distribution is the specificity of these different ions. Cells are rich in potassium and magnesium, and the surrounding plasma is rich in sodium and calcium. It seems likely that the distribution of ions in the plasma of higher animals reflects the oceanic origin of their evolutionary antecedents. One would like to know how primitive cells learned to exclude the sodium and calcium ions in which they were bathed and to develop an internal milieu enriched in potassium and magnesium.

The third and last group of essential elements consists of the trace elements. The fact that they are required in extremely minute quantities in no way diminishes their great importance. In this sense they are comparable to the vitamins. We now know that the great majority of the trace elements, represented by metallic ions, serve chiefly as key components of essential enzyme systems or of proteins with vital functions (such as hemoglobin and myoglobin, which respectively transports oxygen in the blood and stores oxygen in muscle). The heaviest essential element, iodine, is an essential constituent of the thyroid hormones thyroxine and triiodothyronine, although its precise role in hormonal activity is still not understood.

The Trace Elements

To demonstrate that a particular element is essential to life becomes increasingly difficult as one lowers the threshold of the amount of a substance recognizable as a "trace." It has been known for more than 100 years, for example, that iron and iodine are essential to man. In a rapidly developing period of biochemistry between 1928 and 1935 four more elements, all metals, were shown to be

COMPOSITION OF UNIVERSE		COMPOSITION OF EARTH'S CRUST		COMPOSITION OF SEAWATER		COMPOSITION OF HUMAN BODY	
PERCENT OF TOTAL NUMBER OF ATOMS							
H	91	O	47	H	66	H	63
He	9.1	Si	28	O	33	O	25.5
O	.057	Al	7.9	Cl	.33	C	9.5
N	.042	Fe	4.5	Na	.28	N	1.4
C	.021	Ca	3.5	Mg	.033	Ca	.31
Si	.003	Na	2.5	S	.017	P	.22
Ne	.003	K	2.5	Ca	.006	Cl	.03
Mg	.002	Mg	2.2	K	.006	K	.06
Fe	.002	Ti	.46	C	.0014	S	.05
S	.001	H	.22	Br	.0005	Na	.03
		C	.19			Mg	.01
ALL OTHERS	<.01	ALL OTHERS	<.1	ALL OTHERS	<.1	ALL OTHERS	<.01

Al ALUMINUM, B BORON, Br BROMINE, Ca CALCIUM, C CARBON, Cl CHLORINE, He HELIUM, H HYDROGEN, Fe IRON, Mg MAGNESIUM, Ne NEON, N NITROGEN, O OXYGEN, K POTASSIUM, Si SILICON, Na SODIUM, S SULFUR, Ti TITANIUM

CHEMICAL SELECTIVITY OF EVOLUTION can be demonstrated by comparing the composition of the human body with the approximate composition of seawater, the earth's crust and the universe at large. The percentages are based on the total number of atoms in each case; because of rounding the totals do not exactly equal 100. Elements in the colored boxes in the last column appear in one or more columns at the left. Thus one sees that phosphorus, the sixth most plentiful element in the body, is a rare element in inanimate nature. Carbon, the third most plentiful element, is also very scarce elsewhere.

ELEMENT	SYMBOL	ATOMIC NUMBER	COMMENTS
HYDROGEN	H	1	Required for water and organic compounds.
HELIUM	He	2	Inert and unused.
LITHIUM	Li	3	Probably unused.
BERYLLIUM	Be	4	Probably unused; toxic.
BORON	B	5	Essential in some plants; function unknown.
CARBON	C	6	Required for organic compounds.
NITROGEN	N	7	Required for many organic compounds.
OXYGEN	O	8	Required for water and organic compounds.
FLUORINE	F	9	Growth factor in rats; possible constituent of teeth and bone.
NEON	Ne	10	Inert and unused.
SODIUM	Na	11	Principal extracellular cation.
MAGNESIUM	Mg	12	Required for activity of many enzymes; in chlorophyll.
ALUMINUM	Al	13	Essentiality under study.
SILICON	Si	14	Possible structural unit of diatoms; recently shown to be essential in chicks.
PHOSPHORUS	P	15	Essential for biochemical synthesis and energy transfer.
SULFUR	S	16	Required for proteins and other biological compounds.
CHLORINE	Cl	17	Principal cellular and extracellular anion.
ARGON	A	18	Inert and unused.
POTASSIUM	K	19	Principal cellular cation.
CALCIUM	Ca	20	Major component of bone; required for some enzymes.
SCANDIUM	Sc	21	Probably unused.
TITANIUM	Ti	22	Probably unused.
VANADIUM	V	23	Essential in lower plants, certain marine animals and rats.
CHROMIUM	Cr	24	Essential in higher animals; related to action of insulin.
MANGANESE	Mn	25	Required for activity of several enzymes.
IRON	Fe	26	Most important transition metal ion; essential for hemoglobin and many enzymes.
COBALT	Co	27	Required for activity of several enzymes; in vitamin B_{12}.
NICKEL	Ni	28	Essentiality under study.
COPPER	Cu	29	Essential in oxidative and other enzymes and hemocyanin.
ZINC	Zn	30	Required for activity of many enzymes.
GALLIUM	Ga	31	Probably unused.
GERMANIUM	Ge	32	Probably unused.
ARSENIC	As	33	Probably unused; toxic.
SELENIUM	Se	34	Essential for liver function.
MOLYBDENUM	Mo	42	Required for activity of several enzymes.
TIN	Sn	50	Essential in rats; function unknown.
IODINE	I	53	Essential constituent of the thyroid hormones.

SOME TWO-THIRDS OF LIGHTEST ELEMENTS, or 21 out of the first 34 elements in the periodic table, are now known to be essential for animal life. These 21 plus molybdenum (No. 42), tin (No. 50) and iodine (No. 53) constitute the total list of the 24 essential elements, which are here enclosed in colored boxes. It is possible that still other light elements will turn out to be essential. The most likely candidates are aluminum, nickel and germanium. The element boron already appears to be essential for some plants.

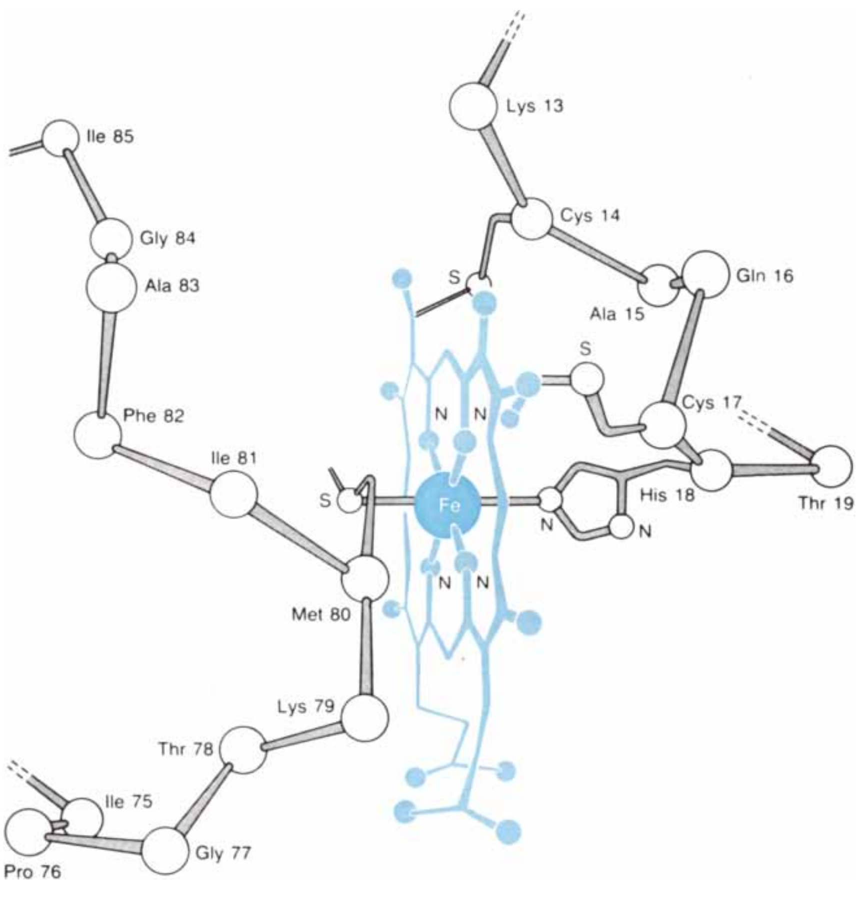
essential: copper, manganese, zinc and cobalt. The demonstration can be credited chiefly to a group of investigators at the University of Wisconsin led by C. A. Elvehjem, E. B. Hart and W. R. Todd. At that time it seemed that these four metals might be the last of the essential trace elements. In the next 30 years, however, three more elements were shown to be essential: chromium, selenium and molybdenum. Fluorine, silicon, tin and vanadium have been added since 1970.

The essentiality of five of these last seven elements was discovered through the careful, painstaking efforts of Klaus Schwarz and his associates, initially located at the National Institutes of Health and now based at the Veterans Administration Hospital in Long Beach, Calif. For the past 15 years Schwarz's group has made a systematic study of the trace-element requirements of rats and other small animals. The animals are maintained from birth in a completely isolated sterile environment [*see illustration on page 59*].

The apparatus is constructed entirely of plastics to eliminate the stray contaminants contained in metal, glass and rubber. Although even plastics may contain some trace elements, they are so tightly bound in the structural lattice of the material that they cannot be leached out or be picked up by an animal even through contact. A typical isolator system houses 32 animals in individual acrylic cages. Highly efficient air filters remove all trace substances that might be present in the dust in the air. Thus the animals' only access to essential nutrients is through their diet. They receive chemically pure amino acids instead of natural proteins, and all other dietary ingredients are screened for metal contaminants.

Since the standards of purity employed in these experiments far exceed those for reagents normally regarded as analytically pure, Schwarz and his coworkers have had to develop many new analytical chemical methods. The most difficult problem turned out to be the purification of salt mixtures. Even the purest commercial reagents were contaminated with traces of metal ions. It was also found that trace elements could be passed from mothers to their offspring. To minimize this source of contamination animals are weaned as quickly as possible, usually from 18 to 20 days after birth.

With these precautions Schwarz and his colleagues have within the past several years been able to produce a new deficiency disease in rats. The animals grow poorly, lose hair and muscle tone, develop shaggy fur and exhibit other detrimental changes [*see illustration on page 60*]. When standard laboratory food is given these animals, they regain their normal appearance. At first it was thought that all the symptoms were caused by the lack of one particular trace element. Eventually four different elements had to be supplied to complete the highly purified diets the animals had been receiving. The four elements proved to be fluorine, silicon, tin and vanadium. A convenient source of these elements is yeast ash or liver preparations from a healthy animal. The animals on the deficiency diet grew less than half as fast as those on a normal or supplemented diet. Growth alone, however, may not tell the entire story. There is some evidence that even the addition of the four elements may not reverse the loss of hair and skin changes resulting from the deficiency diet.



Ala	ALANINE	His	HISTIDINE	Phe	PHENYLALANINE
Cys	CYSTEINE	Ile	ISOLEUCINE	Pro	PROLINE
Gln	GLUTAMINE	Lys	LYSINE	Thr	THREONINE
Gly	GLYCINE	Met	METHIONINE		

THE METALLOENZYME CYTOCHROME *c* is typical of metal-protein complexes in which trace metals play a crucial role. Cytochrome c belongs to a family of enzymes that extract energy from food molecules. It consists of a protein chain of 104 amino acid units attached to a heme group (*color*), a rosette of atoms with an atom of iron at the center. This simplified molecular diagram shows only the heme group and several of the amino acid units closest to it. The iron atom has six coordination sites enabling it to form six bonds with neighboring atoms. Four bonds connect to nitrogen atoms in the heme group itself, and the remaining two bonds link up with amino acid units in the protein chain (histidine at site No. 18 and methionine at site No. 80). The illustration is based on the work of Richard E. Dickerson of the California Institute of Technology, in whose laboratory the complete structure of horse-heart cytochrome *c* was recently determined.

Functions of Trace Elements

The addition of tin and vanadium to the list of essential trace metals brings to 10 the total number of trace metals needed by animals and plants. What role do these metals play? For six of the eight trace metals recognized from earlier studies (that is, for iron, zinc, copper, cobalt, manganese and molybdenum) we are reasonably sure of the answer. The six are constituents of a wide range of enzymes that participate in a variety of metabolic processes [*see illustration at right*].

In addition to its role in hemoglobin and myoglobin, iron appears in succinate dehydrogenase, one of the enzymes needed for the utilization of energy from sugars and starches. Enzymes incorporating zinc help to control the formation of carbon dioxide and the digestion of proteins. Copper is present in more than a dozen enzymes, whose roles range from the utilization of iron to the pigmentation of the skin. Cobalt appears in enzymes involved in the synthesis of DNA and the metabolism of amino acids. Enzymes incorporating manganese are involved in the formation of urea and the metabolism of pyruvate. Enzymes incorporating molybdenum participate in purine metabolism and the utilization of nitrogen.

These six metals belong to a group known as transition elements. They owe their uniqueness to their ability to form strong complexes with ligands, or molecular groups, of the type present in the side chains of proteins. Enzymes in which transition metals are tightly incorporated are called metalloenzymes, since the metal is usually embedded deep inside the structure of the protein. If the metal atom is removed, the protein usually loses its capacity to function as an enzyme. There is also a group of enzymes in which the metal ion is more loosely associated with the protein but is nonetheless essential for the enzyme's activity. Enzymes in this group are known as metal-ion-activated enzymes. In either group the role of the metal ion may be to maintain the proper conformation of the protein, to bind the substrate (the molecule acted on) to the protein or to donate or accept electrons in reactions where the substrate is reduced or oxidized.

In 1968 the complete three-dimensional structure of the first metalloenzyme, cytochrome *c*, was published [see "The Structure and History of an Ancient Protein," by Richard E. Dickerson; SCIENTIFIC AMERICAN, April]. Cytochrome *c*, a red enzyme containing iron, is universally present in plants and animals. It is one of a series of enzymes, all called cytochromes, that extract energy from food molecules by the stepwise addition of oxygen.

METAL	ENZYME	BIOLOGICAL FUNCTION
IRON	FERREDOXIN	Photosynthesis
	SUCCINATE DEHYDROGENASE	Aerobic oxidation of carbohydrates
IRON IN HEME	ALDEHYDE OXIDASE	Aldehyde oxidation
	CYTOCHROMES	Electron transfer
	CATALASE	Protection against hydrogen peroxide
	[HEMOGLOBIN]	Oxygen transport
COPPER	CERULOPLASMIN	Iron utilization
	CYTOCHROME OXIDASE	Principal terminal oxidase
	LYSINE OXIDASE	Elasticity of aortic walls
	TYROSINASE	Skin pigmentation
	PLASTOCYANIN	Photosynthesis
	[HEMOCYANIN]	Oxygen transport in invertebrates
ZINC	CARBONIC ANHYDRASE	CO_2 formation; regulation of acidity
	CARBOXYPEPTIDASE	Protein digestion
	ALCOHOL DEHYDROGENASE	Alcohol metabolism
MANGANESE	ARGINASE	Urea formation
	PYRUVATE CARBOXYLASE	Pyruvate metabolism
COBALT	RIBONUCLEOTIDE REDUCTASE	DNA biosynthesis
	GLUTAMATE MUTASE	Amino acid metabolism
MOLYBDENUM	XANTHINE OXIDASE	Purine metabolism
	NITRATE REDUCTASE	Nitrate utilization
CALCIUM	LIPASES	Lipid digestion
MAGNESIUM	HEXOKINASE	Phosphate transfer

WIDE VARIETY OF METALLOENZYMES is required for the successful functioning of living organisms. Some of the most important are given in this list. The giant oxygen-transporting molecules hemoglobin and hemocyanin are included in the list (in brackets) even though they are not strictly enzymes, that is, they do not act as biological catalysts.

The complete amino acid sequence of cytochrome *c* obtained from the human heart was determined some 10 years ago by a group led by Emil L. Smith of the University of California at Los Angeles and by Emanuel Margoliash of Northwestern University. The iron atom is partially complexed with an intricate organic molecule, protoporphyrin, to form a heme group similar to that in hemoglobin. Of the iron atom's six coordination sites, four are attached to the heme group through nitrogen atoms. The other two sites form bonds with the protein chain; one bond is through a nitrogen atom in the side chain of a histidine unit at site No. 18 in the protein sequence and the other bond is through a sulfur atom in the side chain of a methionine unit at site No. 80 [*see illustration on opposite page*].

Although the cytochrome *c* molecule is complicated, it is one of the simplest of the metalloenzymes. Cytochrome oxidase, probably the single most important enzyme in most cells, since it is responsible for transferring electrons to oxygen to form water, is far more complicated. Each molecule contains about 12 times as many atoms as cytochrome *c*, including two copper atoms and two heme groups, both of which participate in transferring the electrons.

More complicated yet is cysteamine oxygenase, which catalyzes the addition of oxygen to a molecule of cysteamine; it contains one atom each of three different metals: iron, copper and zinc. There are many other combinations of metal ions and unique molecular assemblies. An extreme example is xanthine oxidase, which contains eight iron atoms, two molybdenum atoms and two molecules incorporating riboflavin (one of the B vitamins) in a giant molecule more than 25 times the size of cytochrome *c*.

The metal-containing proteins of another group, the metalloproteins, closely

A CHIEF EXECUTIVE OFFICER SPEAKS HIS MIND. PORTION OF AN INTERVIEW WITH
JOHN O. LOGAN, PRESIDENT, UNIVERSAL OIL PRODUCTS COMPANY, DES PLAINES, ILLINOIS

Q: A UN Report in March said that mankind was in serious jeopardy due to adverse effects on the environment...

A: At the present rate, I believe vast areas of America could be virtually uninhabitable in a few decades.

Q: Ignoring theories, Mr. Logan, what can be done quickly?

A: Our company is one on the leading edge of scores of practical solutions. We have processes to capture particulate matter belching from stacks. To remove sulfur from smoke and recover it as a commodity. To abolish bad odors from food-processing plants. Our people analyze the problem, propose a solution, manufacture the device and then install it in operating condition. We are solving problems for companies like Commonwealth Edison in the Midwest . . . GTE Sylvania . . . Reynolds Aluminum . . . Minnesota Mining and Quaker Oats.

Q: Do cars account for nearly half the pollution in the nation now?

A: Yes—according to the U.S. Environmental Protection Agency. And it's needless. UOP has a converter that, used with unleaded gasoline, can dramatically reduce most exhaust pollutants. We are into such diverse fields as artificial kidneys and waste water clean-up. We have repositioned this company to use our technical prowess in the petroleum area for meaningful solutions in broader areas. We intend to make a contribution to the better life—and make money doing it.

Write to Office of The President for details on:

( ) UOP's stake in the Energy Crisis
( ) UOP's stake in the Environment Crisis
( ) UOP's total Capabilities story

UOP, Universal Oil Products Company
Ten UOP Plaza
Des Plaines, Illinois 60016

UOP
The Answer Company

resemble the metalloenzymes except that they lack an obvious catalytic function. Hemoglobin itself is an example. Others are hemocyanin, the copper-containing blue protein that carries oxygen in many invertebrates, metallothionein, a protein involved in the absorption and storage of zinc, and transferrin, a protein that transports iron in the bloodstream. There may be many more such compounds still unrecognized because their function has escaped detection.

The Newest Essential Elements

Much remains to be learned about the specific biochemical role of the most recently discovered essential elements. In 1957 Schwarz and Calvin M. Foltz, working at the National Institutes of Health, showed that selenium helped to prevent several serious deficiency diseases in different animals, including liver necrosis and muscular dystrophy. Rats were protected against death from liver necrosis by a diet containing one-tenth of a part per million of selenium. Comparably low doses reversed the white muscle disease observed in cattle and sheep that happen to graze in areas where selenium is scarce.

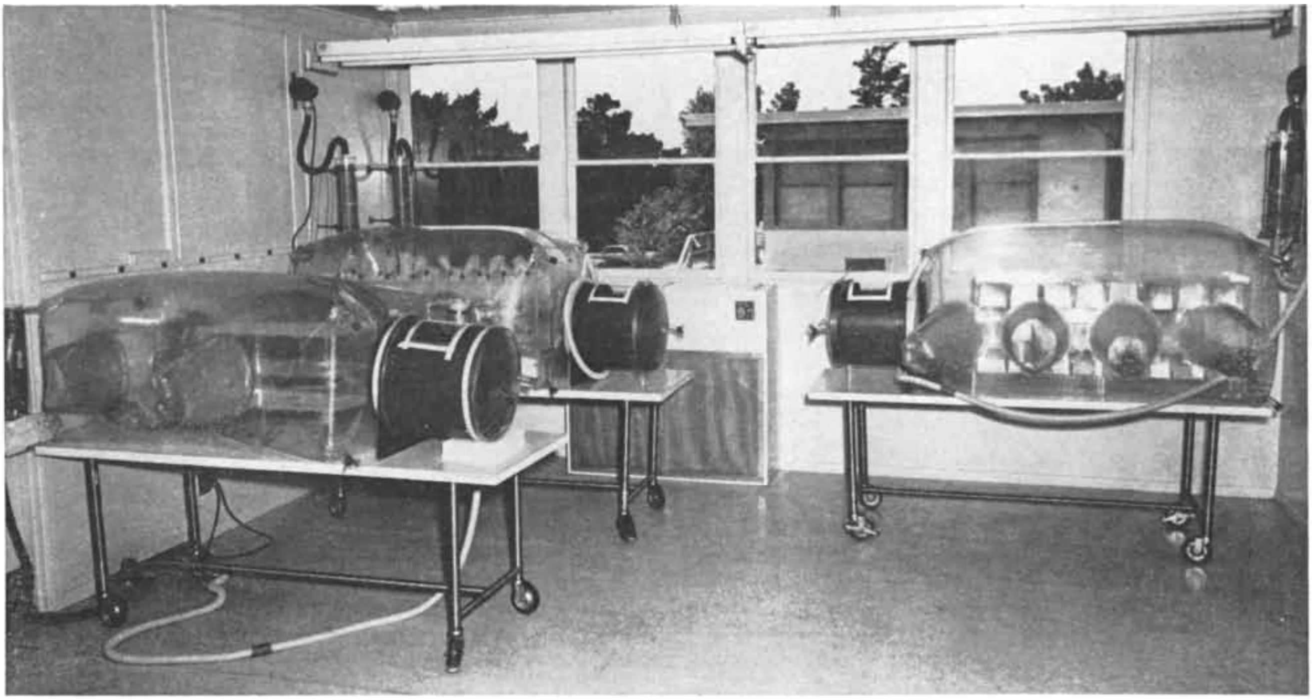
In April a group at the University of Wisconsin under J. T. Rotruck reported a direct biochemical role for selenium. Oxidative damage to red blood cells was detected in rats kept on a selenium-deficient diet. This damage was related to reduced activity of an enzyme, glutathione peroxidase, that helps to protect hemoglobin against the injurious oxidative effects of hydrogen peroxide. The enzyme uses hydrogen peroxide to catalyze the oxidation of glutathione, thus keeping hydrogen peroxide from oxidizing the reduced state of iron in hemoglobin. Oxidized glutathione can readily be converted to reduced glutathione by a variety of intracellular mechanisms. There is some reason to believe glutathione peroxidase may even contain some form of selenium acting as an integral part of the functional enzyme molecule.

The physiological importance of chromium was established in 1959 by Schwarz and Walter Mertz. They found that chromium deficiency is characterized by impaired growth and reduced life-span, corneal lesions and a defect in sugar metabolism. When the diet is deficient in chromium, glucose is removed from the bloodstream only half as fast as it is normally. In rats the deficiency is relieved by a single administration of 20 micrograms of certain trivalent chromic salts. It now appears that the chromium ion works in conjunction with insulin, and that in at least some cases diabetes may reflect faulty chromium metabolism.

After developing the all-plastic trace-element isolator described above, Schwarz, David B. Milne and Elizabeth Vineyard discovered that tin, not previously suspected as being essential, was necessary for normal growth. Without one or two parts per million of tin in their diet, rats grow at only about two-thirds the normal rate.

The next element shown to be essential in mammals by the Schwarz group was vanadium, an element that had been detected earlier in certain marine invertebrates but whose essentiality had not been demonstrated. On a diet in which vanadium is totally excluded rats suffer a retardation of about 30 percent in growth rate. Schwarz and Milne found that normal growth is restored by adding one-tenth of a part per million of vanadium to the diet. At higher concentrations vanadium is known to have several biological effects, but its essential role in trace amounts remains to be established. A high dose of vanadium blocks the synthesis of cholesterol and reduces the amount of phospholipid and cholesterol in the blood. Vanadium also promotes the mineralization of teeth and is effective as a catalyst in the oxidation of many biological substances.

The third element most recently iden-



NUTRITIONAL NEEDS OF SMALL ANIMALS are studied in a trace-element isolator, a modification of the apparatus originally conceived to maintain animals in a germ-free environment. To prevent unwanted introduction of trace elements the isolator is built completely of plastics. It holds 32 animals in separate cages, individually supplied with food of precisely known composition. The system was designed by Klaus Schwarz and J. Cecil Smith of the Veterans Administration Hospital in Long Beach, Calif.

tified as being essential is fluorine. Even with tin and vanadium added to highly purified diets containing all other elements known to be essential, the animals in Schwarz's plastic cages still failed to grow at a normal rate. When up to half a part per million of potassium fluoride was added to the diet, the animals showed a 20 to 30 percent weight gain in four weeks. Although it had appeared that a trace amount of fluorine was essential for building sound teeth, Schwarz's study showed that fluorine's biochemical role was more fundamental than that. In any case fluoridated water provides more than enough fluorine to maintain a normal growth rate.

Although there were earlier clues that silicon might be an essential life element, firm proof of its essentiality, at least in young chicks, was reported only three months ago. Edith M. Carlisle of the School of Public Health at the University of California at Los Angeles finds that chicks kept on a silicon-free diet for only one or two weeks exhibit poor development of feathers and skeleton, including markedly thin leg bones. The addition of 30 parts per million of silicon to the diet increases the chicks' growth more than 35 percent and makes possible normal feathering and skeletal development. Considering that silicon is not only the second most abundant element in the earth's crust but is also similar to carbon in many of its chemical properties, it is hard to see how evolution could have totally excluded it from an essential biochemical role.

Nickel, nearly always associated with iron in natural substances, is another element receiving close attention. Also a transition element, it is particularly difficult to remove from the food used in special diets. Nickel seems to influence the growth of wing and tail feathers in chicks but more consistent data are needed to establish its essentiality. One incidental result of Schwarz's work has been the discovery of a previously unrecognized organic compound, which will undoubtedly prove to be a new vitamin.

TRACE-ELEMENT DEFICIENCY developed when the rat at the top of this photograph was kept in the trace-element isolator for 20 days and fed a diet from which fluorine, tin and vanadium had been carefully excluded. The healthy animal at the bottom was fed the same diet but was kept under ordinary conditions. It was evidently able to obtain the necessary trace amounts of fluorine, tin and vanadium from dust and other contaminants.

Synergism and Antagonism

The interaction of the various essential metals can be extremely complicated. The absence of one metal in the diet can profoundly influence, either positively or negatively, the utilization of another metal that may be present. For example, it has been known for nearly 50 years that copper is essential for the proper metabolism of iron. An animal deprived of copper but not iron develops anemia because the biosynthetic machinery fails to incorporate iron in hemoglobin molecules. It has only recently been found in our laboratories at Florida State University that ceruloplasmin, the copper-containing protein of the blood, is a direct molecular link between the two metals. Ceruloplasmin promotes the release of iron from animal liver so that the iron-binding protein of the serum, transferrin, can complex with iron and transfer it to the developing red blood cells for direct utilization in the biosynthesis of hemoglobin. This represents a synergistic relation between copper and iron.

As an example of antagonism between elements one can cite the instance of copper and zinc. The ability of sheep or cattle to absorb copper is greatly reduced if too much zinc or molybdenum is present in their diet. Evidently either of the two metals can displace copper in an absorption process that probably involves competition for sites on a metal-binding protein in the intestines and liver.

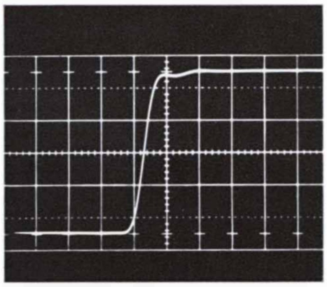
The recent discoveries present many fresh challenges to biochemists. One can expect the discovery of previously unsuspected metalloenzymes containing vanadium, tin, chromium and selenium. New compounds or enzyme systems requiring fluorine and silicon may also be uncovered. The multiple and complex interdependencies of the elements suggest many hitherto unrecognized and important facts about the role and interrelations of metal ions in nutrition and in health and disease.

Every man needs one retreat from mediocrity.

Mercedes-Benz
Mercedes-Benz of North America, Inc.

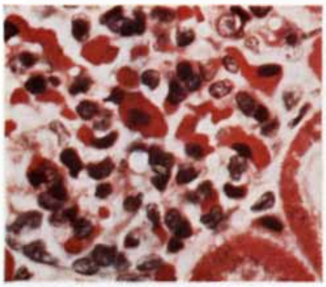
Recording wonders in seconds is one of the wonders.

CR-9 FOR OSCILLOGRAPHY

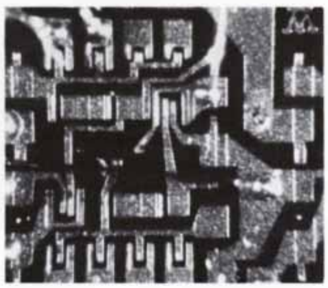


Hard copy, oscilloscope readout

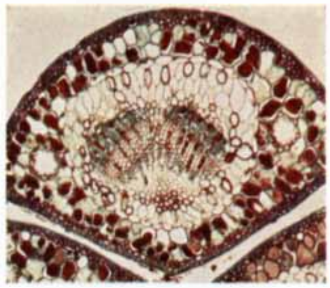
ED-10 FOR PHOTOMICROSCOPY



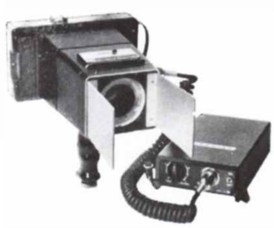
Section, mouse kidney, 1000x



Integrated circuit, 80x

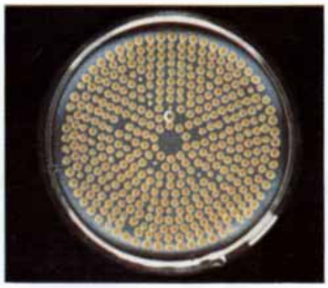


Pinus leaf section, 100x

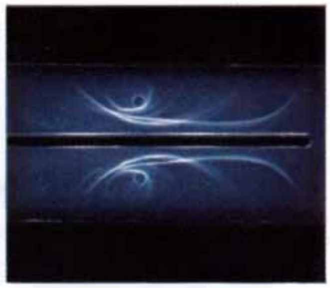


CU-5 FOR CLOSE-UPS

Clinical dental record



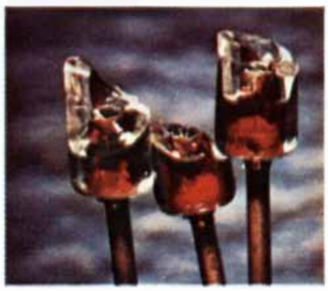
Macrophotograph, bacteria colonies



Immunoelectrophoresis

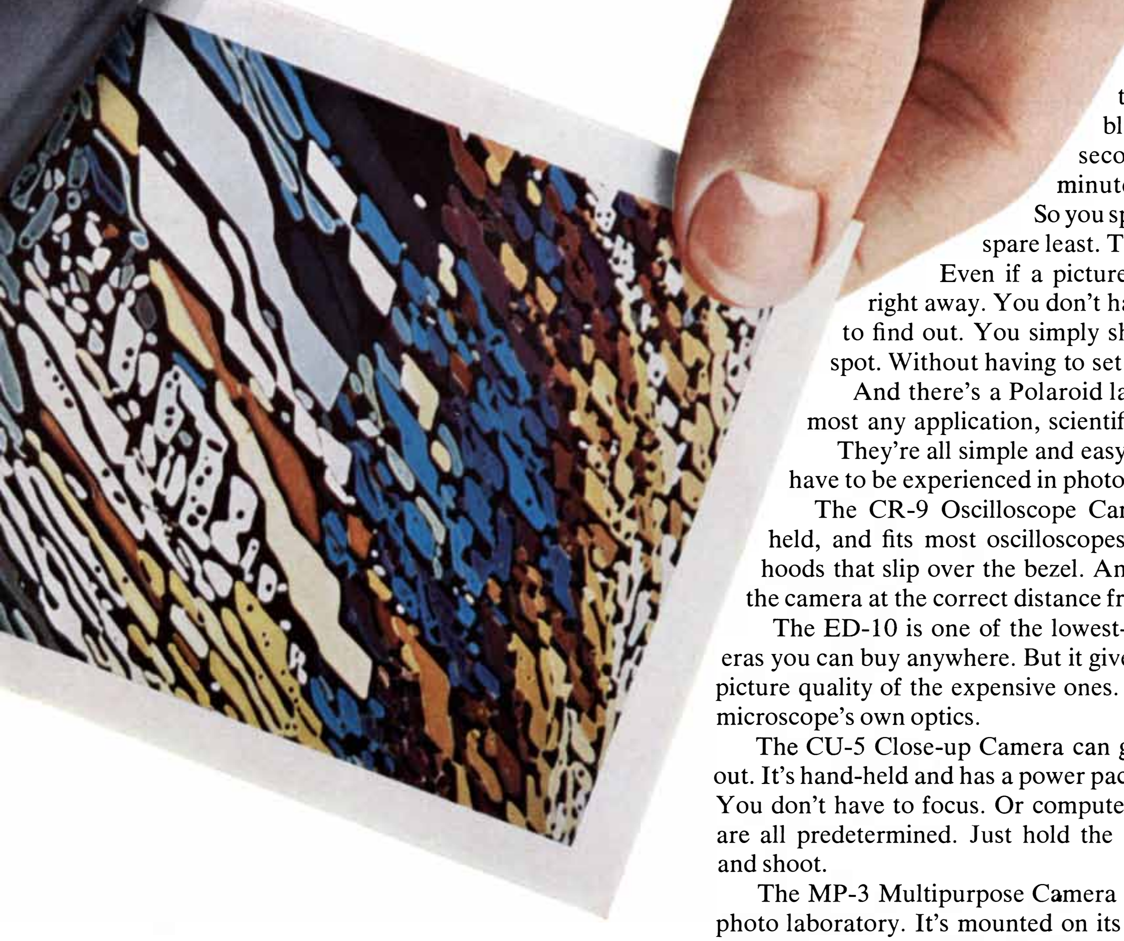
MP-3 MULTIPURPOSE

High contrast copy, continuous-tone art



Macrophotograph, diode fractures

Fingerprint record, negative



What's wonderful is that you get a finished black-and-white print in seconds, or color in just one minute. Of almost any subject. So you spend less on what you can spare least. Time.

Even if a picture gets spoiled, you see it right away. You don't have to wait hours or days to find out. You simply shoot again. Right on the spot. Without having to set things up all over again.

And there's a Polaroid laboratory camera for almost any application, scientific, medical or industrial. They're all simple and easy to operate, so you don't have to be experienced in photography to use them.

The CR-9 Oscilloscope Camera is low cost, hand-held, and fits most oscilloscopes. It has interchangeable hoods that slip over the bezel. And automatically position the camera at the correct distance from the display.

The ED-10 is one of the lowest-priced microscope cameras you can buy anywhere. But it gives you the resolution and picture quality of the expensive ones. Because they all use the microscope's own optics.

The CU-5 Close-up Camera can go anywhere. Indoors or out. It's hand-held and has a power pack to supply its own light. You don't have to focus. Or compute exposure settings. They are all predetermined. Just hold the camera over the subject and shoot.

The MP-3 Multipurpose Camera is really a self-contained photo laboratory. It's mounted on its own stand, with built-in lighting. It uses 13 Polaroid self-developing films, including color; black-and-white; high speed; high contrast; line and continuous tone transparency films; even one that gives you both a positive and a matching negative in seconds. It has 5 lens and shutter combinations. And 3 changeable backs for a variety of film formats. The MP-3 is so versatile, even Polaroid hasn't found out all it can do.

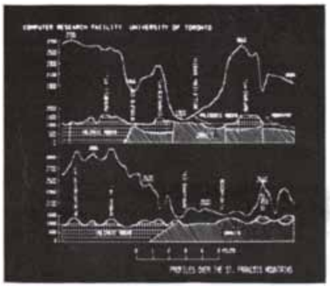
But we know that the Polaroid system of instant photography does what nothing else in the world can. It gives you a permanent record of nearly any subject in seconds. And gives it to you simply and easily.

For full information, or a demonstration of any Polaroid Lab camera, write to Polaroid Corporation, Dept. 57-209, 549 Technology Square, Cambridge, Mass. 02139. In Puerto Rico, P.O. Box 2032, Ceramica Annex, Carolina 00630.

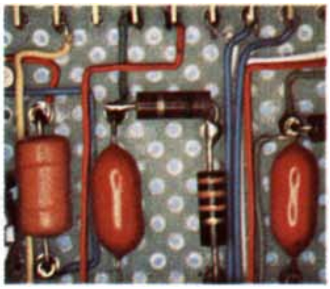
Convincing you of the benefits of the system should be easy. All you need is a little exposure.

Polaroid®

Polaroid Instant Lab Cameras.

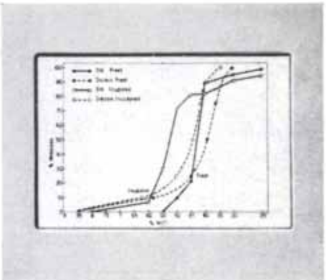


Hard copy, computer terminal display

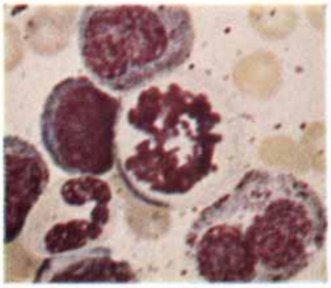


Electronic components

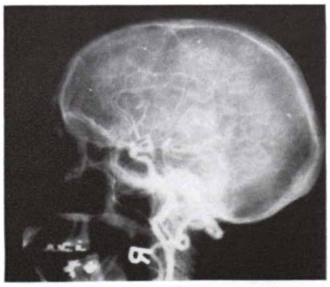
Gross specimen, intestinal fibroma



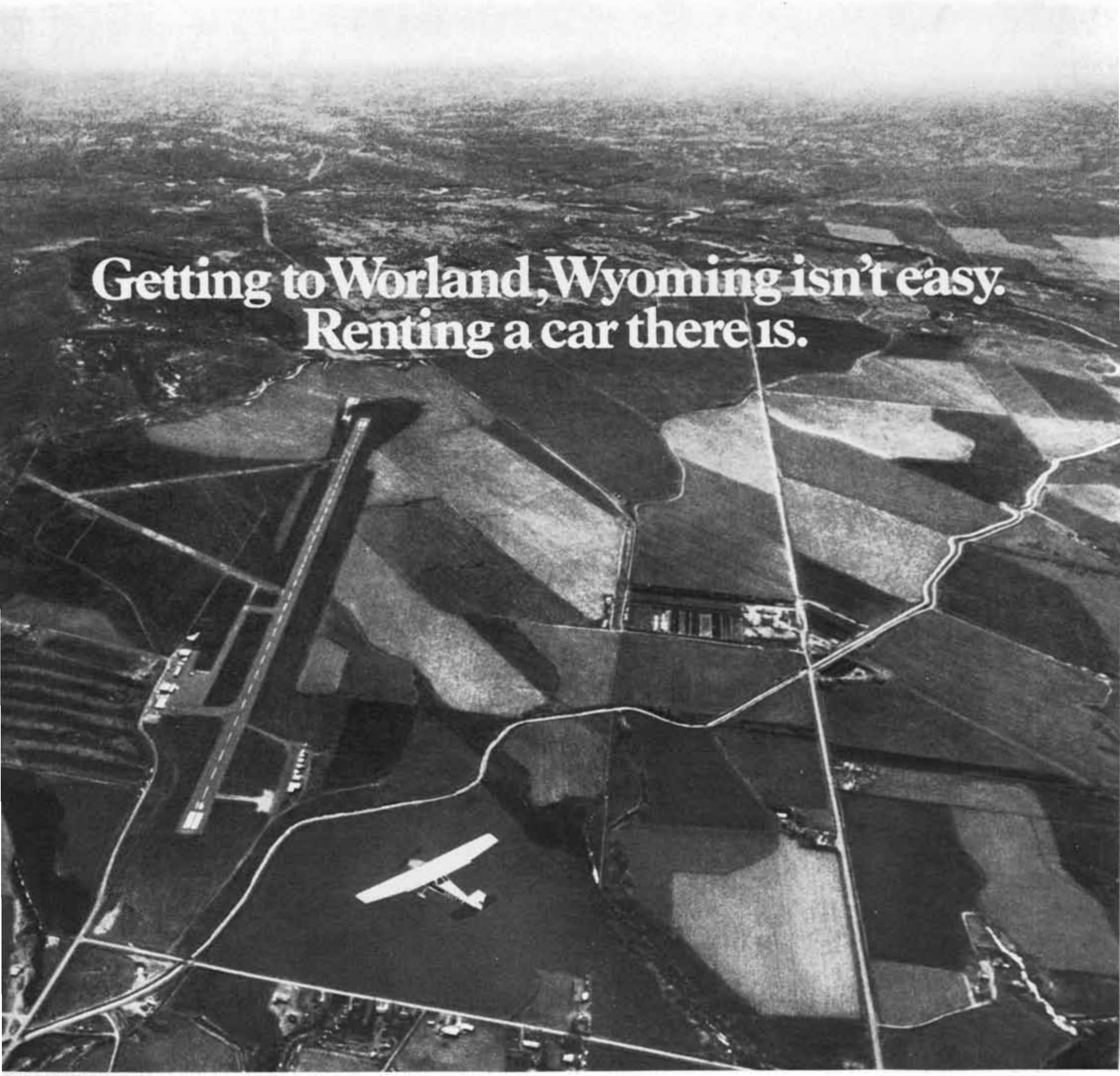
Graph for presentation, 35mm



Human leukemia, 1600x



X-ray copy

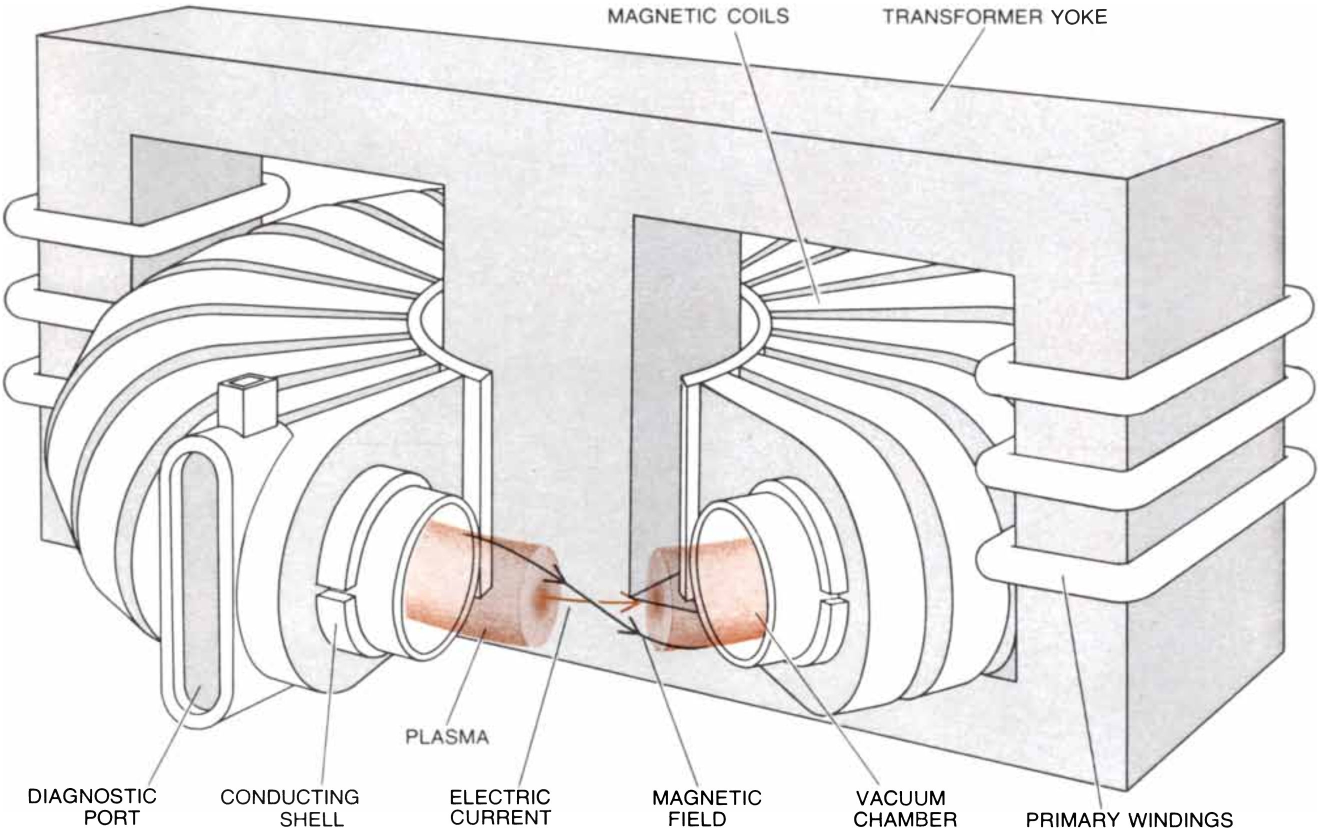

Getting to Worland, Wyoming isn't easy.
Renting a car there is.

The Tokamak Approach in Fusion Research

Experimental studies of plasma heating and confinement in machines based on the Tokamak design, a "toroidal diffuse pinch" configuration, are being conducted and planned in laboratories around the world

by Bruno Coppi and Jan Rem

The resurgence of interest in the prospects of getting useful power from controlled thermonuclear, or fusion, reactions has been stimulated largely by two concurrent developments. One is the growing awareness on the part of a concerned public that most energy sources have severe deficiencies, either in terms of their limited size or in terms of their potentially harmful environmental effects. Equally important are reports of significant progress in experimental efforts to contain a hot plasma of the type needed for fusion reactions. These results have encouraged the hope that a proof of the feasibility of a thermonuclear reactor may be attained in the not too distant future [see "The Prospects of Fusion Power," by William C. Gough and Bernard J. Eastlund; SCIENTIFIC AMERICAN, February, 1971].

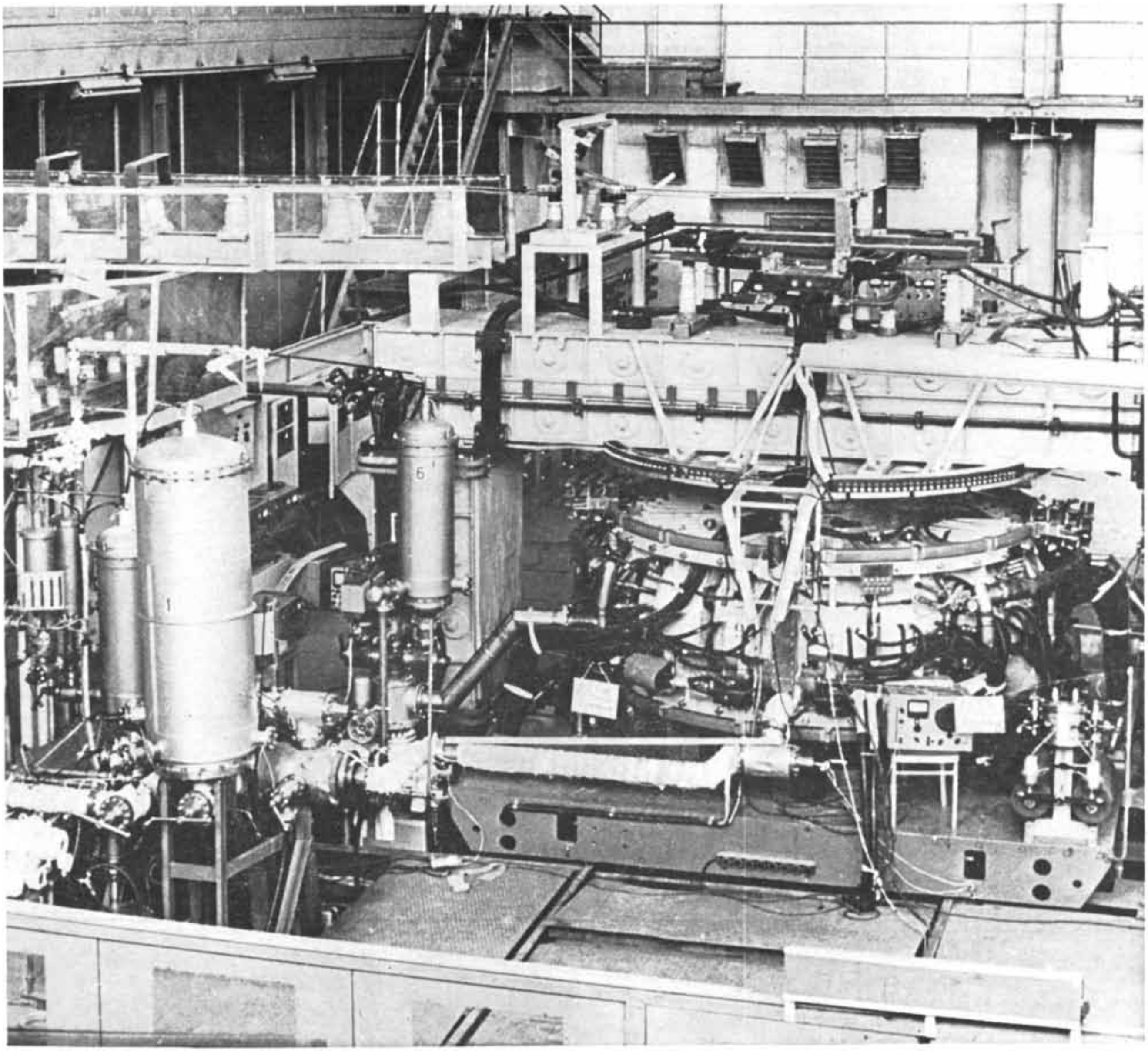


TYPICAL TOKAMAK is depicted in this schematic diagram. The hot plasma, or gas of charged particles (*color*), is confined on a nested complex of magnetic surfaces composed of helical magnetic-field lines, only one of which is indicated (*black arrows*). The helical magnetic field is the resultant of two component fields (*not shown here*), which are oriented at right angles to each other. The toroidal component is set up by the circular array of wedge-shaped coils distributed around the toroidal plasma chamber. The poloidal component is set up by a toroidal current (*colored arrow*) that flows inside the plasma; this toroidal current is in turn created by a toroidal electric field produced by a transformer consisting of a set of primary windings around an iron yoke. (The plasma itself constitutes the secondary winding.) A plasma configuration of this type, in which a doughnut-shaped plasma is subject to a constricting force generated by the current flowing in it, is called a toroidal pinch; a toroidal diffuse pinch is one in which the current is distributed over the cross section of the plasma. The conducting shell keeps the plasma column from expanding in the direction of the major radius. The diagnostic port (*lower left*) provides access for measuring various characteristics of the plasma column.

One of the most active areas of research on controlled fusion involves a class of experimental machines designed to confine a toroidal, or doughnut-shaped, plasma. Within this class the particular configuration that has attracted the most attention in recent years is the "diffuse pinch," a configuration on which the series of Tokamak experiments, developed at the I. V. Kurchatov Institute of Atomic Energy in Moscow, has been based. Since 1968, when the most important results obtained by the Tokamak devices began to be widely accepted within the scientific community, a variety of experiments on the heating and confinement of high-temperature plasmas have been undertaken around the world adopting the toroidal-diffuse-pinch configuration. The machines involved are often referred to generically as Tokamaks. Here we shall describe some of the ideas behind several of these experiments. We shall present not only some of the reasons for confidence in their outcome but also some of the problems that remain to be solved before this approach can provide an experimental feasibility proof of a thermonuclear reactor.

In our opinion the strongest reason for confidence is that since the mid-1960's considerable progress has been made in understanding the dynamics of high-temperature plasmas. One indication of this fact is that at about that time theoretical predictions and experimental findings in plasma physics started to go hand in hand. Before then most of the large-scale experiments were proposed and undertaken on the basis of intuition with not too solid a theoretical foundation or as a continuation of a preexisting line of development for a given laboratory. Meanwhile the theory of plasmas was proceeding on its own, more with a flavor of mathematical physics than with a strong motivation to explain and predict experimental observations. In this light the success of the Tokamak experiments can be seen as one of several important steps, most of them less well known, that have marked the evolution of plasma physics in the past decade.

Although the recent progress of fusion research has been limited by the lack of



LARGEST TOKAMAK currently in operation is the T-4 device, located at the I. V. Kurchatov Institute of Atomic Energy in Moscow. The plasma is formed inside a toroidal chamber with a major radius of 100 centimeters and a minor radius of 20 centimeters.

adequate financial support, this factor has been partly compensated for by the extensive exchange of information and the collaboration on an international scale among institutions engaged in such work.

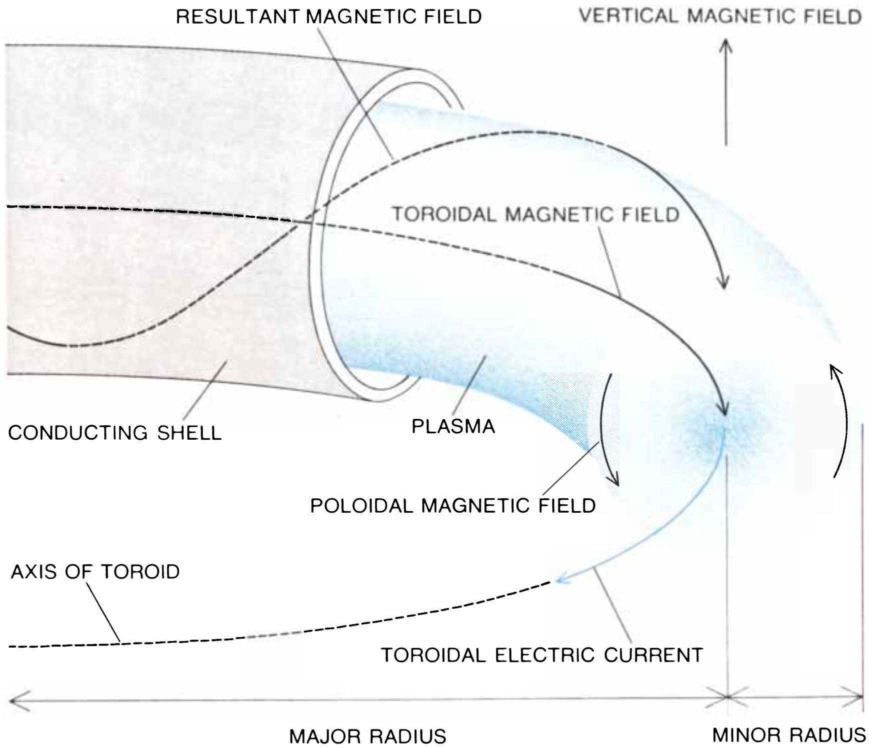
The basic process of nuclear fusion is well known. If two light atomic nuclei are brought together with enough force to overcome the repelling Coulomb force, they fuse, yielding a heavier nucleus and at least one other particle (a proton or a neutron, depending on the reaction); the reaction products are all characterized by extremely high energy. For example, one of the possible fusion reactions involves two isotopes of hydrogen (deuterium and tritium), which when they fuse yield an alpha particle (a helium-4 nucleus) and a neutron with an energy gain of 17.58 million electron volts (MeV).

A fusion reactor could consist simply of a container holding a mixture of fully ionized deuterium and tritium nuclei at a very high temperature. In such a hot plasma fusion reactions would occur when the ignition temperature is reached; at this temperature the energy released by the fusion reactions equals the energy lost by radiation from the plasma. Even for the deuterium-tritium reaction, which has the lowest ignition temperature, this temperature is still quite high: 46 million degrees Kelvin. Clearly a plasma in which thermonuclear reactions will begin cannot be contained by material walls. Most of the envisioned fusion reactors are therefore based on a plasma confined by a magnetic field.

Another criterion that must be met in order to achieve a net production of thermonuclear energy from fusion reactions is that the product of the plasma particle density times the containment time must exceed a certain value: 10^{14} seconds per cubic centimeter. This requirement, called the Lawson criterion after the British physicist J. D. Lawson, corresponds to the condition that during the containment time the released fusion energy, at a temperature higher than the ignition temperature, should at least equal the sum of the energy lost by radiation plus the energy needed to raise the thermal energy of the plasma to the considered temperature.

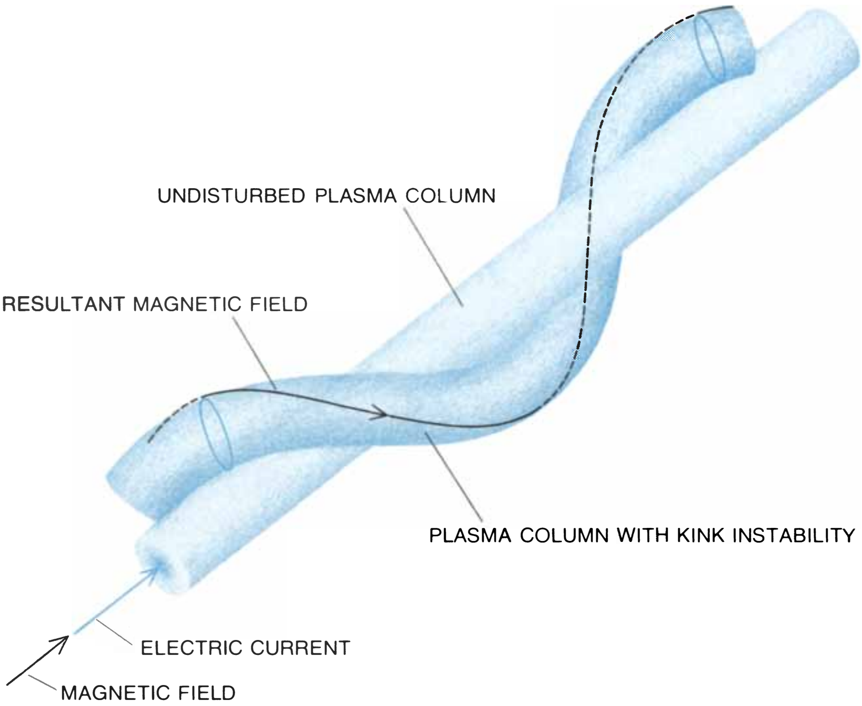
A confined plasma can diffuse through a magnetic field in two ways: as a result of collisions between particles or as a result of instabilities. At the high temperatures typical of fusion plasmas the collision frequencies are so low that the containment time dictated by collisional diffusion is long enough to meet the Lawson criterion. The many instabilities to which a plasma in a magnetic field is susceptible, however, can cause particles and energy to be lost across the magnetic field much faster than the loss rates attributable to collisions between particles. These plasma instabilities can be roughly grouped in two classes: macroscopic and microscopic. Macroscopic instabilities involve large-scale motions of the plasma that tend to destroy its initial configuration. Some of them are quite similar to the hydrodynamic instability that develops when a fluid of given density is superimposed on one of lower density in a downward gravitational field [see "The Leakage Problem in Fusion Reactors," by Francis F. Chen; SCIENTIFIC AMERICAN, July, 1967]. Microinstabilities can lead to fluctuations in density and electric field within the plasma and to the transport of particles and energy across the magnetic field at a higher rate than would be expected from the effects of particle collisions alone. Considerable progress has been made in understanding the nature of such instabilities and in finding ways to prevent the most damaging among them.

In the past decade a large variety of magnetic-field configurations have been devised and investigated for their containment and stability properties. Although we shall discuss only one type of toroidal confinement scheme—the toroidal diffuse pinch—many of the observations we shall make apply just as well to other types of configuration.

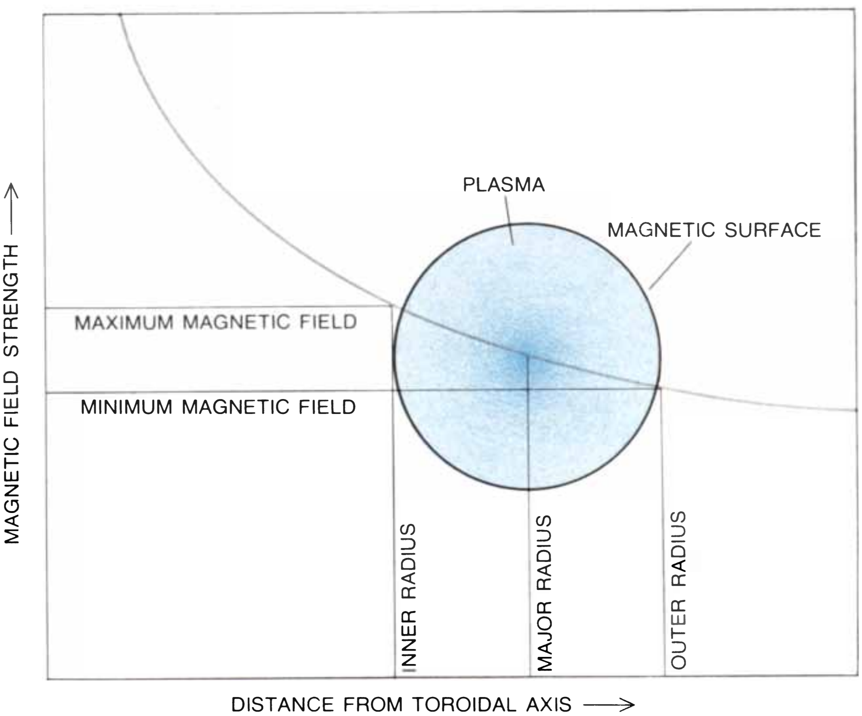


MAJOR PARAMETERS of a toroidal-diffuse-pinch experiment are indicated in this cutaway view. It is the combination of the toroidal magnetic field set up by the external coils and the poloidal magnetic field set up by the electric current flowing in the plasma that produces the helical field lines shown in the illustration on page 65. Expansion of the plasma column in the direction of the minor radius is prevented by the pinch force generated by the current; expansion in the direction of the major radius is prevented either by the conducting metallic shell surrounding the plasma or by the vertical magnetic field.

The toroidal diffuse pinch is one of the simplest schemes for the magnetic containment of high-temperature plasmas. It consists of a plasma ring with a circular cross section embedded in a strong toroidal magnetic field; a toroidal electric field maintains a toroidal current that flows inside the plasma, and this current in turn generates a magnetic-field component that is poloidal [*see illustration above*]. The combination of the poloidal field with the toroidal field produces helical magnetic-field lines that lie on closed magnetic surfaces (nested toroids of circular cross section). The plasma configurations referred to generically as pinches are characterized by having the plasma subject to a constricting force generated by the current flowing in them. The pinching force per unit volume is equal to the strength of the poloidal magnetic field times the current density (that is, the current per unit



"KINK" INSTABILITY is one of the most serious macroscopic instabilities that can disrupt a current-carrying plasma column in a longitudinal magnetic field. The kink instability arises when the current is so strong that over the length of the plasma column a magnetic-field line (*black arrow*) circles the column just once. The growing helical displacement that the column then undergoes has the same pitch as the magnetic-field line.

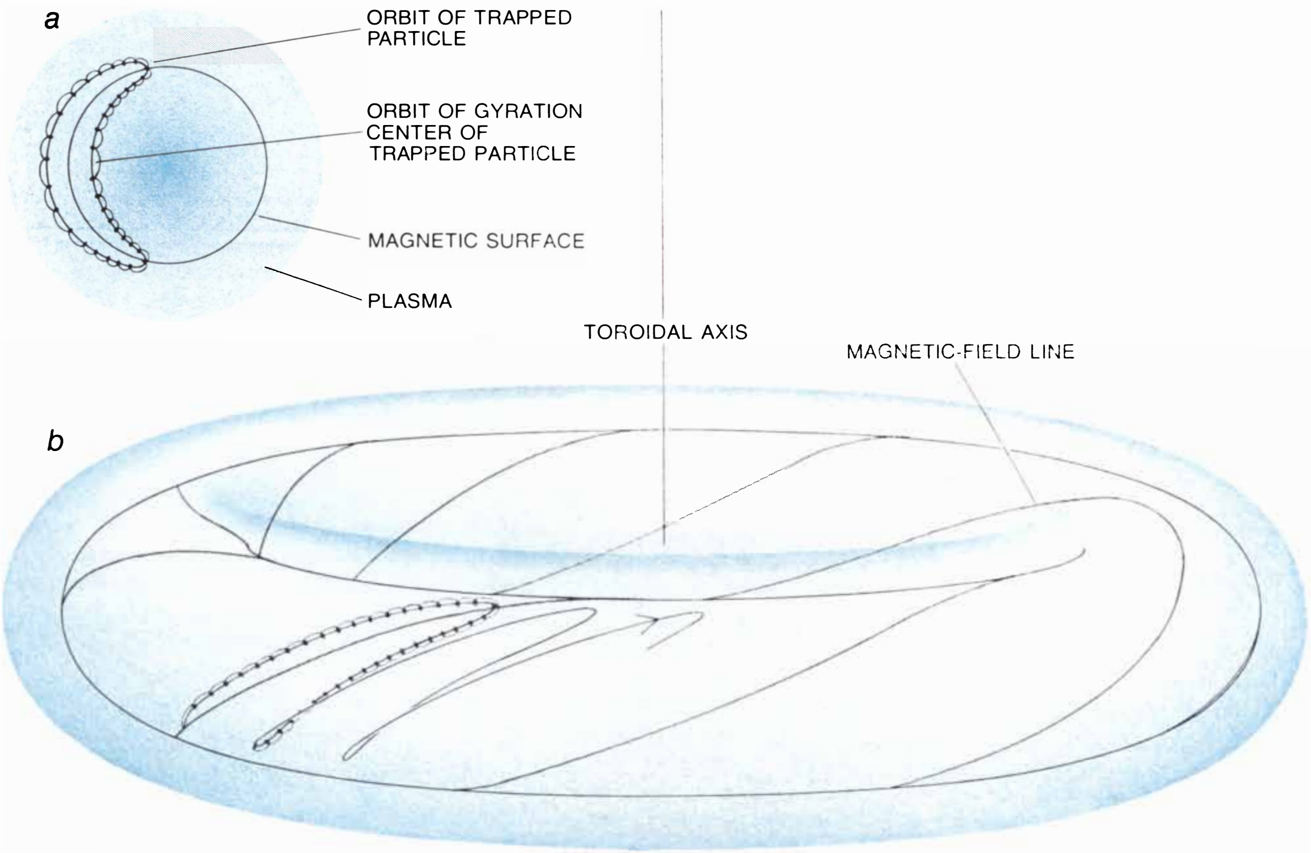


MAGNETIC-FIELD STRENGTH along a helical field line in a toroidal plasma column fluctuates because the field varies inversely with distance from the toroidal axis (*gray curve*). Hence in following such a field line one encounters a maximum and a minimum in the magnetic-field strength every time one circles the magnetic surface (*black circle*). This variation causes particles with low velocities parallel to the field line to be trapped and to oscillate about the minimum-strength regions (*see illustration on opposite page*).

area of the cross section). The same type of force is responsible for the attraction between two parallel current-carrying wires if the currents are in the same direction. A "sharp" pinch is one in which the current along the plasma column flows within a thin surface layer. Conversely, a "diffuse" pinch is one in which the current is distributed over the cross section of the plasma. The main components of an experimental machine for generating a toroidal diffuse pinch are a toroidal plasma chamber, a coil to provide the toroidal magnetic field and a transformer to induce the electric field (the plasma itself constitutes the secondary winding). The plasma chamber is in effect a vacuum chamber because the particle density typical of such experiments (10^{13} to 10^{14} particles per cubic centimeter) is very low compared with that of a gas at atmospheric pressure and temperature.

A plasma without a current is not in equilibrium in a toroidal magnetic field; it is like a toroidal balloon without surface tension that tends to expand in all directions. In the presence of a toroidal current equilibrium becomes possible; the expansion in the direction of the minor radius is then prevented by the pinch force and the expansion in the direction of the major radius is restricted by the effects of a conducting shell surrounding the plasma or by a force associated with a properly imposed vertical magnetic field. Without a vertical magnetic field the latter expansion causes the poloidal magnetic field to be compressed between the plasma and the conducting shell; on the short time scale characteristic of a plasma-confinement experiment the field cannot penetrate the plasma or the shell because both are good conductors. The compressed poloidal magnetic field exerts a force on the plasma that limits the expansion in the direction of the major radius. Even when such a shell is sufficient to prevent the expansion, a vertical field is still applied to directly control the position of the plasma column.

The most serious macroscopic instability to which a toroidal diffuse pinch is susceptible is the "kink" instability, which arises when the current in the plasma exceeds a certain value known as the Kruskal-Shafranov limit (after Martin D. Kruskal of Princeton University and the Russian plasma physicist V. D. Shafranov). At this limit the poloidal magnetic field is such that a magnetic-field line fits just once on the length of the plasma column and once around it, closing on itself. The column is then



"BANANA REGIME" is a term used to describe the situation in which the confined particles in a toroidal-diffuse-pinch configuration are separated into two classes: trapped particles that oscillate around the region of minimum magnetic-field strength and circulating particles that are free to move along a field line for its entire length. The center of gyration of a typical trapped particle has an orbit that is characterized by an oscillatory motion between two reflection points along a given magnetic-field line and an upward (or downward) motion due to the variation of the toroidal magnetic field in the direction of the major radius. The resulting orbit when projected on a cross section of the magnetic surface has a characteristic banana shape (*a*). The corresponding projection for the orbit of a typical circulating particle would coincide almost exactly with the cross section of the magnetic surface. The combination of the oscillatory motion of the trapped particle and its upward (or downward) drift gives rise to a drift along the torus (*b*).

subjected to a growing helical displacement that destroys the equilibrium configuration [*see top illustration on opposite page*]. Expressed in somewhat different terms, this instability arises when the ratio of the toroidal magnetic-field strength to the poloidal one equals the ratio of the major radius to the minor radius (a value known as the aspect ratio of the toroid). Since the poloidal magnetic-field strength is proportional to the toroidal current, this limit implies that for a given aspect ratio the kink instability sets in at higher values of the current with a higher strength of toroidal magnetic field.

In addition to the kink instability there are other types of macroscopic instability in diffuse pinches. For reasonable distributions of the current inside the plasma column, however, such instabilities can be avoided if the toroidal current is kept below the Kruskal-Shafranov limit by a safety factor that is usually taken to be between 2.5 and 3. For practical reasons the aspect ratio is always smaller than 1/3, and so this condition implies that the poloidal magnetic field is considerably weaker than the toroidal magnetic field.

All macroscopic instabilities can be described by a theory that treats the plasma as a conducting fluid and ignores its particle aspect. This point of view is no longer adequate at the high temperatures typical of thermonuclear plasmas, where the mean free paths for interparticle collisions become longer than the length of the system. To illustrate one aspect of the problem we shall consider the motion of a charged particle in the magnetic field of a toroidal diffuse pinch. As we have shown, the combination of the toroidal and the poloidal fields results in helical magnetic-field lines that lie on a toroidal surface. Along such a line the magnetic-field strength varies because the toroidal component of the field varies inversely with distance from the toroidal axis, so that in following the field line one encounters a maximum and a minimum in the magnetic-field strength every time one circles the magnetic surface [*see bottom illustration on opposite page*]. This variation in the magnetic-field strength causes particles with low velocities parallel to the field line to be trapped about the minimum-strength regions; those with a large enough parallel velocity are not trapped and move on a field line for its entire length. These circulating particles carry the current that is induced by the externally applied electric field.

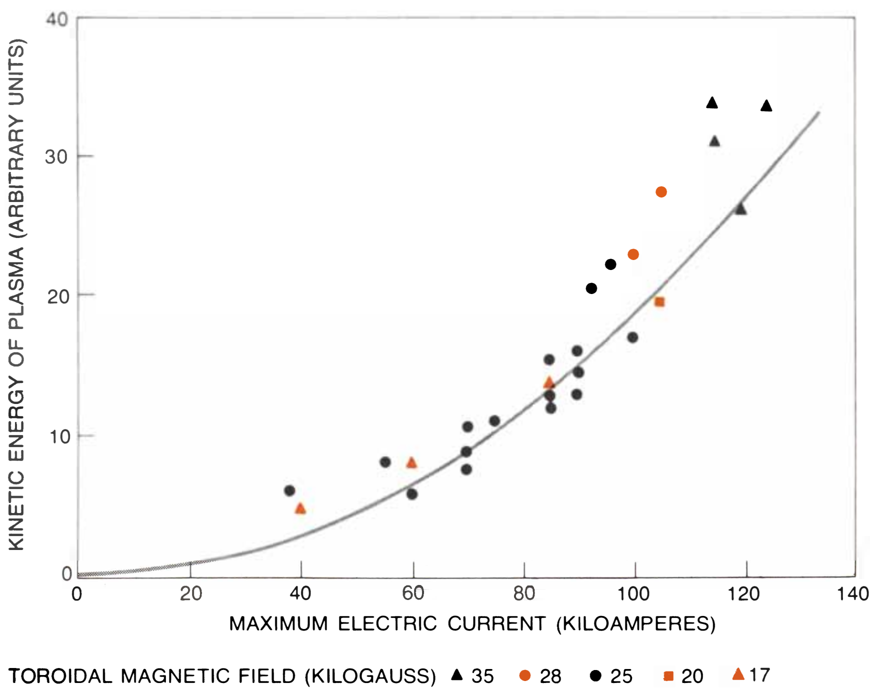
The distinction between circulating particles and trapped particles is well represented by the projection of the center of gyration of a particle on a cross section of the plasma [*see illustration above*]. In the case of a circulating particle this cross section is a full circle slightly displaced from the magnetic surface; in the case of a trapped particle it is a complex banana-like curve. Accordingly the situation where there is such a separation of particles into these two classes is called the "banana regime." In this regime there can be new instabilities and new forms of particle transport and energy transport across the magnetic field. One worrisome aspect of this situation is the theoretical prediction that the trapped particle can leave the system by an appropriate instability mechanism. The experiments carried out so far have arrived just at the border line of the banana regime, and no clear evidence of such an instability has appeared as yet.

Theoretical analysis based only on the effects of particle collisions shows that the diffusion of particles and energy across the magnetic field does not depend on the toroidal field but rather on the poloidal field, that is, on the field created by the current. Therefore the principal function of the toroidal field is to prevent the macroinstabilities mentioned above, and raising its value can only have a favorable effect on the plasma confinement to the extent that the poloidal field (and therefore the current) can be increased. This general feature of toroidal-confinement configurations has been supported by the experimental evidence gathered so far.

The possibility of producing plasmas of thermonuclear interest in a diffuse-pinch configuration became apparent through the encouraging results obtained from experiments carried out in the 1960's on a series of Tokamaks at the Kurchatov Institute. A typical Tokamak device consists of a toroidal vacuum chamber surrounded by a set of coils (to provide a strong toroidal magnetic field) and linked to an iron-core transformer. A varying current in the primary circuit of this transformer induces through the changing magnetic flux in the iron yoke an electric field in the toroid, which in turn drives the current in the plasma. The plasma current will run only as long as the magnetic flux is changing. Since the maximum available flux is limited by the cross section of the iron yoke and by the maximum magnetic field it can handle before saturation, the experiment is necessarily time-limited. For various reasons the current should not fill the entire cross section of the vacuum chamber. To restrict the plasma to a given radius a "limiter" is placed inside the toroid. This device often consists of a plate with a circular hole in its center; the plate of course must be made of a material with a high melting point.

The most advanced Tokamak currently in operation is the T-4 device [*see illustration on page 66*]. In this machine the plasma is formed inside a stainless-steel chamber with a major radius of 100 centimeters and a minor radius of 20 centimeters. At the strongest possible toroidal magnetic field of 40 kilogauss the stability criterion given above would allow a plasma column with a minor radius of 17 centimeters to carry a current of 300 kiloamperes. The best plasma parameters obtained so far correspond to electron temperatures of 2,000 to 3,000 electron volts, ion temperatures of about 600 electron volts, densities of about 3×10^{13} particles per cubic centimeter and energy-confinement times of some 20 milliseconds. (A temperature of 1,000 electron volts is equivalent to 11.6 million degrees K.)

Among the many important conclusions that have been reached with the Russian Tokamak experiments, one is that the maximum thermal energy density at which a plasma can be confined is roughly equal to the magnetic energy density associated with the poloidal magnetic field. Since this magnetic energy density is proportional to the square of the poloidal magnetic field, and since the poloidal field is proportional to the toroidal current, the measured kinetic energy density is expected to vary with the square of the current. This relation was confirmed in 1969 by a joint British-Russian team; by measuring the scattering of laser light from the plasma electrons, they were able to determine accurately both the electron density and the temperature [*see illustration at left*]. Implicit in this result is the fact that the energy-containment time is longer than the time required for resistive heating of the plasma up to the maximum thermal energy density indicated above.



IMPORTANT EXPERIMENTAL RESULTS of the Russian Tokamak experiments confirmed that the maximum kinetic energy density of a plasma confined in a toroidal diffuse pinch is proportional to the square of the maximum current carried by the plasma. This relation is clearly indicated by this graph, in which the plotted experimental points correspond to different values of the toroidal magnetic field (*see key at bottom*). The plasma density and electron temperature were determined by a joint British-Russian team at the Kurchatov Institute by measuring the scattering of laser light from the plasma electrons.

The simplest form of heating that can be adopted in toroidal plasma confinement experiments is ohmic, or resistive, heating associated with the flow of current inside the plasma. Hence in a toroidal diffuse pinch the toroidal current can provide confinement and heating at the same time (a feature of all Tokamak experiments performed so far).

The resistivity of a plasma in its simplest form results from electron-ion collisions. For an electric field applied along the toroidal magnetic field the electrons will tend to gain momentum in the direction opposite to the electric field and the ions in the direction of the electric field. Electron-ion collisions make the electrons lose momentum to the ions, however, and this prevents them from being freely accelerated, thus producing a resistivity. This collisional resistivity, sometimes referred to as "classical" resistivity, is a strongly decreasing inverse

EXPERIMENT		MAJOR RADIUS (CENTIMETERS)	MINOR RADIUS (CENTIMETERS)	TOROIDAL MAGNETIC-FIELD STRENGTH (KILOGAUSS)	ESTIMATED MAXIMUM PLASMA CURRENT (KILOAMPERES)
FRANCE	TFR	98	20	60	400
GERMANY	PULSATOR	70	13	28	135*
ITALY	FRASCATI	82	22 — 24	100	1,150*
JAPAN	JFT-2	90	25	15	250
U.S.	DOUBLET II (GULF)	59	30 — 90	10	350
	ALCATOR (M.I.T.)	54	11.5	120	600*
	ORMAK (OAK RIDGE)	79.5	23.4	25	400
	ST (PRINCETON)	109	14	35	100
	ATC (PRINCETON)	30	7	60	270
	PLT (PRINCETON)	130	45	50	1,600
	TTT (TEXAS)	60	10	35	100
U.S.S.R.	TOKAMAK TM-3	40	8	25	80*
	TOKAMAK T-4	100	17	40	300
	TOKAMAK T-6	70	25	15	268*
	TOKAMAK T-10	150	40	40	850*

VARIETY OF EXPERIMENTS based on the Tokamak concept is suggested by this table, which lists all the toroidal-diffuse-pinch machines now in existence or in the design stage in laboratories around the world. Numbers give the major parameters of the experiments. The numbers followed by a superscript star were obtained using a safety factor of 2.5 against macroscopic instabilities.

function of the electron temperature. Therefore if only electron-ion collisions are taken into consideration, a plasma with an electron temperature of 1,000 electron volts would be about as resistive as copper at room temperature.

A result one would expect from the temperature dependence of the classical resistivity and of the collisional electron thermal conductivity across the magnetic field is the formation of a "skin current." Calculations based on these transport coefficients show that the current density tends to acquire a pronounced maximum near the edge of the plasma, the same type of profile that is obtained for the electron temperature. In all experiments, however, the current is found to be a smoothly decreasing function of the radius for almost the entire duration of the discharge. The conclusion is clear: The current distribution is not controlled by collisional effects.

The thermal power generated by the current in the plasma is proportional to the resistance and to the square of the current density. Since the current density is limited by the stability criterion mentioned above, if the plasma resistivity is merely equal to the collisional resistivity, the input power will strongly decrease when the plasma temperature increases. Therefore to obtain higher temperatures than those yet achieved it will be important to devise heating methods with a more favorable temperature dependence than the heating that results from collisional resistivity.

It is known that when certain types of wave are excited inside the plasma, there can be a transfer of momentum from the electrons to the wave and from the wave to the ions. In this way a resistivity is produced without involving single-particle collisions. When this form of resistivity prevails over the collisional form, it is usually termed "anomalous resistivity." If the current density is kept below the limit for the onset of macroinstabilities but above the threshold value for the onset of microinstabilities, the plasma tends to be slightly turbulent but not so macroscopically unstable that its confinement is lost. Therefore a promising method of heating high-temperature plasmas is to enhance the resistivity over its collisional values by allowing for the excitation of waves by current-driven microinstabilities in the plasma itself.

This type of heating will be employed in at least one of the new experiments to be described below. Since the threshold value increases with temperature, the current density should be high for this heating scheme to remain effective at high temperatures. For that reason the design parameters of such a device must be chosen so that the current density allowed by the stability limit is as high as possible.

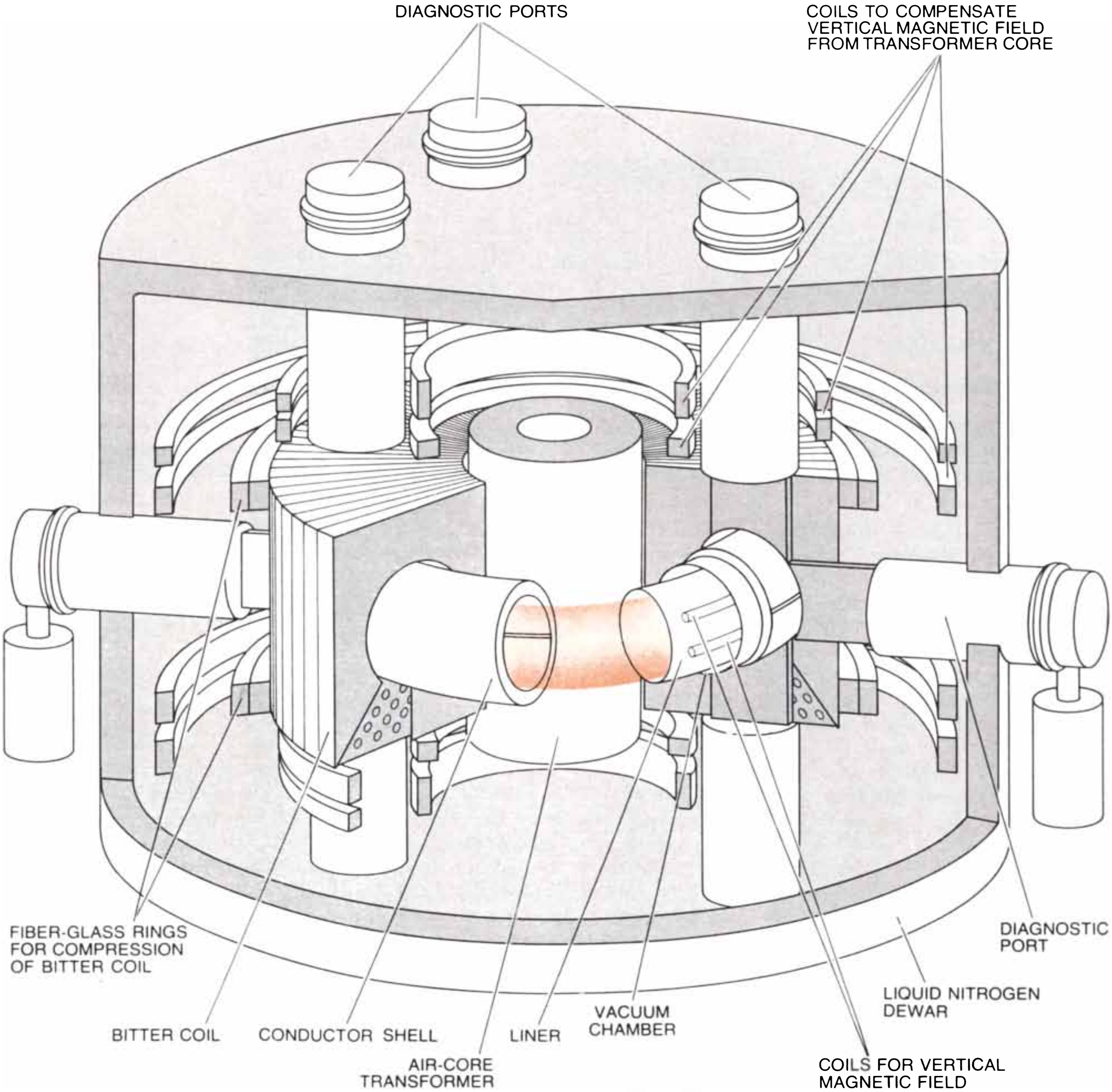
In addition, experiments are planned in order to see whether or not it is possible to enhance the plasma microturbulence at high temperatures by the injection of electromagnetic waves with appropriate frequencies so as to resonate with the plasma's own waves. With ohmic heating, either by collisional or by anomalous resistivity, the energy source is essentially the magnetic energy stored in the transformer.

Another turbulent heating scheme consists in exciting a strong plasma turbulence for a very short time by applying a short but strong electric-field pulse. This induces a large current density, which in turn gives rise to plasma instabilities. It is hoped that the plasma will not be lost during this pulse, and that its temperature will be raised and then maintained at very high values.

A completely different approach is to heat a plasma column by compressing it. One possibility is to attempt to reduce the column's major and minor radii simultaneously. A related heating scheme proposed in 1969 by Lev A. Artsimovich of the Kurchatov Institute consists in oscillating the plasma column in the direction of the major axis. In both schemes the heating results from the transformation of magnetic energy into thermal energy.

Still another approach to complementing ohmic heating at high temperatures is to inject a high-energy beam of neutral, or un-ionized, atoms. The injected neutral atoms penetrate the magnetic field and then become ionized, thereby sharing their energy with the ions of the background plasma.

The technical simplicity of the toroidal-diffuse-pinch configuration allows for a variety of experiments aimed at attaining quite different plasma conditions. This has been one of the important factors leading to the design and construction of several machines based on the Tokamak concept in large national laboratories around the world. Most of the new devices are in fact complementary in their aims. To illustrate this complementarity we shall mention a few of the more important devices now operating or under construction.



ALCATOR, a toroidal-diffuse-pinch machine currently under construction at the Massachusetts Institute of Technology, is portrayed in this cutaway drawing. This particular version of the basic Tokamak principle is designed to produce and confine a plasma with properties similar to those of a thermonuclear plasma. The plasma will be heated by inducing within it a large current density, which is expected to generate microturbulence and hence an enhanced electrical resistivity. A toroidal plasma column that carries such a large current, however, tends to be destroyed by macroscopic fluid-like instabilities unless a large enough toroidal magnetic field is applied. Accordingly the Alcator will be built to attain toroidal magnetic fields up to about 120 kilogauss in a plasma chamber with a major radius of 54 centimeters and a minor radius of 12.5 centimeters. This extremely strong magnetic field should enable the plasma column in Alcator to carry currents up to 600 kiloamperes with a reasonable safety factor against macroscopic instabilities.

The first proposed experiment to obtain more advanced plasma parameters than those realized in the Russian Tokamak experiments by 1969 is the Alcator device, which is being constructed at the Massachusetts Institute of Technology [*see illustration on opposite page*]. This machine is designed to create confined plasmas with densities and temperatures that are characteristic of the banana regime and are therefore close to thermonuclear plasmas.

To reach high-temperature regimes in relatively dense plasmas it is important to have a high specific heating rate. Therefore the parameters of Alcator have been chosen so that the plasma is allowed to carry very high current densities. The heating rate is expected to be enhanced by the fact that for the contemplated high current densities the resistivity should be larger than would result from particle collisions at temperatures of up to 5,000 electron volts and plasma densities of about 10^{14} particles per cubic centimeter. Since a large value of the current density implies a large value of the poloidal magnetic field, the maximum kinetic energy of the plasma that can be confined is also high.

As we have mentioned, the current density must remain below the Kruskal-Shafranov limit by a certain safety factor in order to prevent macroinstabilities from destroying the confinement. This limit is proportional to the toroidal magnetic-field strength and inversely proportional to the major radius. Hence, in order to have a large current density and to be able to contain a plasma with a large kinetic energy (which is proportional to the square of the product of the minor radius and the current density), the toroidal magnetic field and the minor radius should be as large as possible and the major radius should be as small as possible. In practice these are conflicting requirements. Nonetheless, the application of advanced techniques developed at M.I.T.'s National Magnet Laboratory, where Alcator is being built under the direction of D. B. Montgomery, should make it possible to attain toroidal magnetic fields of up to 120 kilogauss in a plasma chamber with a major radius of 54 centimeters and a minor radius of 12.5 centimeters. The toroidal magnetic field is created by a current flowing in a toroidal Bitter coil (named for the late Francis Bitter of M.I.T.) made of copper and reinforced by stainless-steel plates. The coil is kept at the temperature of liquid nitrogen (77 degrees K.) in order to decrease its resistivity. (At this temperature the resistivity of copper is about a seventh of what it is at room temperature.) Therefore with the available voltage a higher current can be driven through the coil and higher values of the toroidal magnetic field can be reached. For this a power of 24 megawatts is provided by the four generators of the laboratory; it is expected that the maximum magnetic field will be maintained for a quarter of a second.

The current needed to heat the plasma and to contain it is induced by means of an air-core transformer located at the center of the Bitter coil. It can be shown that for the confined plasma to reach the densities and temperatures for which Alcator is designed the transformer must generate a magnetic flux of at least 1.2 volt-seconds. Unlike the transformers of the existing Tokamaks, the transformer of Alcator does not have an iron yoke, since the magnetic field of a transformer that has to fit into the relatively small space at the center of the Bitter coil and still deliver a flux of about 1.2 volt-seconds must be roughly 120 kilogauss, which is far higher than the saturation threshold of iron (the maximum field strength that can be carried by iron).

The criterion for macroscopic stability with a safety factor of 2.5 allows a plasma column with a minor radius of 11.5 centimeters to carry a current of about 600 kiloamperes. The corresponding current density of 1,400 amperes per square centimeter is almost five times larger than that achieved in the best Tokamak experiments. Therefore for equal plasma conditions the specific heating rate of Alcator could in principle be about 25 times higher. The pressure of the plasma that can be contained with this current would correspond to a particle density of 2.5×10^{14} particles per cubic centimeter and a temperature of about 5,000 electron volts. The form of heating that has been adopted, however, is expected to lead to these temperatures only with smaller particle densities.

To ensure equilibrium the plasma column is surrounded by a thin copper shell cooled with liquid nitrogen. The poloidal magnetic field cannot penetrate the copper shell, so that the plasma column is kept from moving outward for a time of up to a quarter of a second. The copper shell is mounted on a vacuum chamber provided with four ports in order to allow for diagnostic access to the plasma. The measurements to be carried out consist mostly in analyzing the radiation coming from the plasma, studying the transmitted or reflected radiation from outside sources and evaluating the electric and magnetic fields created by the plasma currents.

It is important that the plasma obtained be made up of pure hydrogen or deuterium ions. A small quantity of heavier atoms can cause a substantial increase in radiation losses and can limit the attainable temperature. In addition, the collisional resistivity can be increased by the presence of impurities and can obscure the effects of low-level turbulence. In particular, impurities freed from the wall of the vacuum chamber by collisions with energetic ions can enter the plasma. Heating the wall before the experiment would partly remove these impurities and so decrease the effect. Therefore a bakable toroidal chamber called a liner is placed between the plasma and the vacuum chamber.

The high energy densities that should be achieved in the Alcator experiments present for the first time in fusion research practical problems such as how to terminate the plasma-confinement phase without damaging the device. Thus in addition to the "scientific" risks associated with the attempt to attain new types of plasma regime there are many technological risks as well.

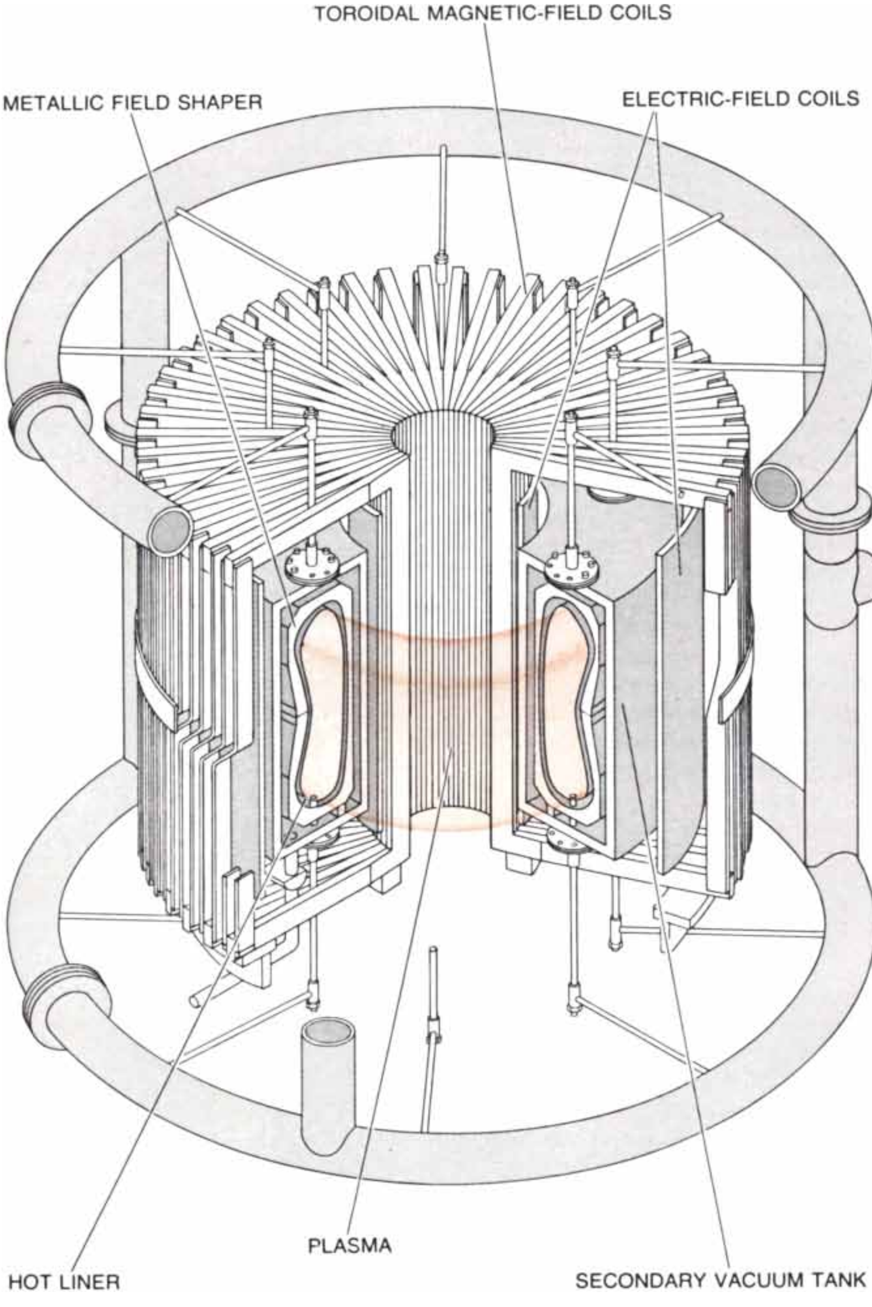
Some problems, such as the expected copious emission of X rays from a high-temperature plasma, which appreciably affects the overall energy balance in the plasma, can turn out to be profitable from a different point of view. For example, an effort will be made to correlate the X-ray emission from Alcator with the emission of known X-ray stars, for which the major plasma parameters are quite well determined [see "The X-Ray Sky," by Herbert W. Schnopper and John P. Delvaille, page 26].

The first large Tokamak machine put into operation in the U.S. resulted from a conversion of the Model *C* stellarator in the Princeton University Plasma Physics Laboratory. The converted machine, designated Model ST, was initially aimed at reproducing the main results of the Russian experiments. The higher repetition rate of the Princeton device and the application of advanced diagnostic techniques, however, have yielded a wealth of new information that not only has confirmed the results obtained in the Russian machines but also has improved considerably the understanding of the dynamics of the contained plasmas.

A large toroidal-diffuse-pinch machine named Ormak has been built at the Oak Ridge National Laboratory. This device, which has now completed its preliminary phase of operation, is aimed at studying plasma confinement in a toroid with a low aspect ratio and a relatively large volume (in other words, a fat toroid). In Ormak it will be possible for the first time to test plasma heating by the injection of high-energy neutral atoms.

If a large electric field is applied to a plasma, instabilities can arise that result in strong turbulence and produce a large anomolous resistivity of the plasma. In relatively simple experiments this has been shown to lead to a rapid heating of the plasma. These experiments, however, were not designed for an effective confinement of the plasmas produced. Accordingly a toroidal-diffuse-pinch experiment recently undertaken at the University of Texas is aimed at exploring how well a plasma that is heated to a high temperature in a short but strongly turbulent phase can be contained. The device has been named the Texas Turbulent Tokamak (TTT).

All toroidal-diffuse-pinch configurations built or designed so far are expected to produce plasmas with a low ratio of thermal energy density to the magnetic energy density of the field in which they are confined. This ratio is usually termed the beta ratio. A thermonuclear plasma should not have too small a beta ratio or radiation losses become too large. It has been shown theoretically that the maximum beta ratio of a plasma that can be confined in a toroidal diffuse pinch can be increased by the proper shaping of the plasma cross section. This possibility will be investigated in a device called Doublet, which has been built by Gulf Energy and Environmental Systems [*see illustration at left*].



DOUBLET II is a diffuse-pinch device with a noncircular cross section built by Gulf Energy and Environmental Systems. The aim of this machine will be to contain plasmas with higher energy densities than those that can be contained in a configuration with a circular cross section. The plasma's cross section will be defined by a metallic field-shaper.

Another experiment at Princeton is aimed at first creating a plasma with parameters similar to those of the plasmas obtained in the Model ST experiment and heating it further by compression [*see illustration on opposite page*]. The experiment, called ATC (for adiabatic toroidal compression), also allows for heating by the injection of high-energy neutral atoms and by microwaves.

The most advanced Russian Tokamak planned, the T-10, is a large machine conceived as an evolutionary continuation of the earlier Tokamak experiments. It is designed to test the influence of large plasma cross sections on containment time and ion temperature. Tentative proposals for an experiment of this kind were discussed as early as 1966. It is estimated that the final design will have a minor radius of 40 centimeters, a major radius of 150 centimeters and a toroidal magnetic field of 40 kilogauss. An advanced machine along the same lines is planned for Princeton; it is designated PLT (for Princeton large torus). The PLT device is expected to have a minor radius of 45 centimeters, a major radius of 130 centimeters and a toroidal magnetic field of 50 kilogauss. PLT is designed to allow for a variety of heating schemes and large diagnostic ports on the confined plasma.

We have noted that the maximum current density that can be carried by the plasma in a diffuse pinch is proportional to the ratio of the toroidal magnetic field to the major radius of the plasma. Thus the current density and hence the heating rate in experiments with large plasma dimensions are necessarily low. A parameter that is simultaneously a measure of the current density and the confinement time that can be attained in a device is the total current. Confinement time can in fact be shown to depend strongly on the cross-sectional area of the plasma. Accordingly an experiment aimed at producing plasma currents on the order of one megampere without sacrificing the current density will be constructed at the laboratory at Frascati near Rome as part of a cooperative effort sponsored by Euratom. The major parameters are a minor radius of 22 centimeters, a major radius of 83 centimeters and a toroidal magnetic field of 100 kilogauss. If, as is often done, the product of the toroidal magnetic field times the plasma radius is taken as a parameter of merit to gauge the performance of a device, the PLT and the Frascati machines have the common goal of 2,200 kilogauss-centimeter. In the table on page 71 we list all the Tokamak-like devices that are in operation or being constructed around the world for a variety of purposes.

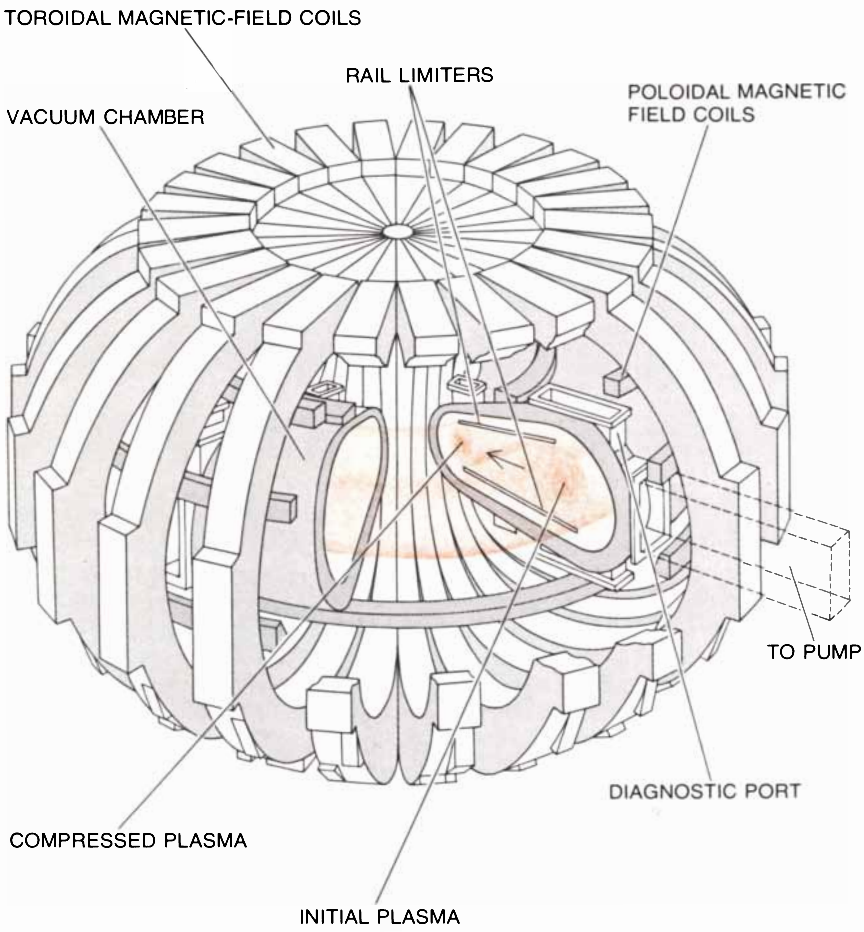
There are a number of unknowns associated with plasma confinement in the regimes that should be attained by the new types of toroidal diffuse pinch now coming into operation. For example, when the temperature increases, the role of particle collisions in determining the plasma transport parameters (electrical resistivity, particle diffusion and heat diffusion) tends to become weak. A low level of turbulence created by the excitation of plasma microinstabilities can then considerably alter the value of these parameters with respect to those expected on the basis of theories that consider only the effects of particle collisions. Because of its inherent difficulties the theory of transport in turbulent plasmas (for instance of turbulent resistivity) is still full of uncertainties. Thus a prediction of the outcome of the new generation of experiments at high-temperature regimes is quite difficult.

In high-temperature plasmas, which are characterized by low collision frequencies between particles, the particle population separates into two classes: the magnetically trapped particles and the circulating particles. A theoretical analysis indicates that in this case a large fraction of trapped particles can be subject to instabilities and lose their confinement. On the other hand, there are theoretical arguments indicating that other kinds of microinstability can be present simultaneously and can provide a "mixing" of trapped and circulating particles. If the mixing occurs at a fast enough rate in comparison with the rate of development of the instabilities involving only trapped particles, these instabilities can be avoided.

In spite of the large electrical conductivity of the plasmas produced in diffuse pinches, these plasmas have not exhibited an appreciable "skin" effect preventing the penetration of current inside the plasma column. Collective effects, such as instabilities arising when the electron temperature increases toward the outer edge of the plasma column, are in all probability responsible for the phenomenon. It is not certain, however, that this favorable circumstance will persist through temperature regimes higher than those achieved so far.

As for how an economic thermonuclear fusion reactor might be based on the toroidal-diffuse-pinch principle, there remain several difficulties. One is connected with the fact that losses by synchrotron radiation (radiation emitted by energetic electrons gyrating in the strong toroidal magnetic field) call for a ratio of the kinetic energy density of the particles to the magnetic energy density not smaller than .1. Present experiments show that the values of this ratio that are compatible with confinement are quite low. The reason for this limitation is not well understood, but it is not difficult to anticipate that a number of configurations other than a diffuse pinch with a circular cross section will be considered in order to raise the limiting value of the thermal energy density. Another difficulty is that a toroidal diffuse pinch, at least in its present mode of operation, is inherently a pulsed device because it needs an induced electric field to drive a constant toroidal current, and in a thermonuclear reactor steady-state operation is highly desirable. Toward that end there are theoretical proposals for a steady-state diffuse pinch, where the confining current (called the bootstrap current) is a by-product of the diffusion of trapped particles across the magnetic field.

In our opinion the solution of such problems belongs to a later phase of the thermonuclear research program. For the present it is important to know how dense, high-temperature plasmas can be confined for a significant length of time regardless of the strength of the toroidal magnetic field in which the confinement is achieved, and to gain a reasonable understanding of the dynamics of the process. For this goal the Tokamak approach appears to offer the best promise.



ATC EXPERIMENT, built at the Princeton University Plasma Physics Laboratory, is aimed at first creating a hot toroidal plasma and then heating it further by compression. (The name ATC stands for adiabatic toroidal compression.) This experiment will also test heating schemes based on the injection of high-energy neutral atoms and of microwaves.

DEPRIVATION DWARFISM

Children raised in an emotionally deprived environment can become stunted. The reason may be that abnormal patterns of sleep inhibit the secretion of pituitary hormones, including the growth hormone

by Lytt I. Gardner

It has long been known that infants will not thrive if their mothers are hostile to them or even merely indifferent. This knowledge is the grain of truth in a tale, which otherwise is surely apocryphal, told about Frederick II, the 13th-century ruler of Sicily. It is said that Frederick, himself the master of six languages, believed that all men were born with an innate language, and he wondered what particular ancient tongue—perhaps Hebrew—the language was. He sought the answer through an experiment. A group of foster-mothers was gathered and given charge of certain newborn infants. Frederick ordered the children raised in silence, so that they would not hear one spoken word. He reasoned that their first words, owing nothing to their upbringing, would reveal the natural language of man. "But he labored in vain," the chronicler declares, "because the children all died. For they could not live without the petting and the joyful faces and loving words of their foster mothers."

Similar emotional deprivation in infancy is probably the underlying cause of the spectacularly high mortality rates in 18th- and 19th-century foundling homes [see "Checks on Population Growth: 1750–1850," by William L. Langer; SCIENTIFIC AMERICAN, February]. This, at least, was the verdict of one Spanish churchman, who wrote in 1760: "In the foundling home the child becomes sad, and many of them die of sorrow." Disease and undernourishment certainly contributed to the foundlings' poor rate of survival, but as recently as 1915 James H. M. Knox, Jr., of the Johns Hopkins Hospital noted that, in spite of adequate physical care, 90 percent of the infants in Baltimore orphanages and foundling homes died within a year of admission.

Only in the past 30 years or so have the consequences of emotional deprivation in childhood been investigated in ways that give some hope of understanding the causative mechanisms. One pioneer in the field, Harry Bakwin of New York University, began in 1942 to record the physiological changes apparent in infants removed from the home environment for hospital care. These children, he noted, soon became listless, apathetic and depressed. Their bowel movements were more frequent, and even though their nutritional intake was adequate, they failed to gain weight at the normal rate. Respiratory infections and fevers of unknown origin persisted. All such abnormalities, however, quickly disappeared when the infants were returned to their home and mother.

Another early worker, Margaret A. Ribble, studied infants at three New York maternity hospitals over a period of eight years; in several instances she was able to follow the same child from birth to preadolescence. When normal contact between mother and infant was disrupted, she noted, diarrhea was more prevalent and muscle tone decreased. The infants would frequently spit up their food and then swallow it again. This action is called "rumination" by pediatricians. Largely as the result of studies by Renata and Eugenio Gaddini of the University of Rome, it is recognized today as a symptom of psychic disturbance. Ribble concluded that alarm over a lack of adequate "mothering" was not mere sentimentality. The absence of normal mother-infant interaction was "an actual privation which may result in biological, as well as psychological, damage to the infant."

Two other psychiatrically oriented investigators of this period, René Spitz of the New York Psychoanalytic Institute and his colleague Katherine Wolf, took histories of 91 foundling-home infants in the eastern U.S. and Canada. They found that the infants consistently showed evidence of anxiety and sadness. Their physical development was retarded and they failed to gain weight normally or even lost weight. Periods of protracted insomnia alternated with periods of stupor. Of the 91, Spitz and Wolf reported, 34 died "in spite of good food and meticulous medical care." The period between the seventh and the 12th month of life was the time of the highest fatalities. Infants who managed to survive their first year uniformly showed severe physical retardation.

The comparative irrelevance of good diet, as contrasted with a hostile environment, was documented with startling clarity in Germany after World War II. In 1948 the British nutritionist Elsie M. Widdowson was stationed with an army medical unit in a town in the British Zone of Occupation where two small municipal orphanages were located. Each housed 50-odd boys and girls between four and 14 years of age; the children's average age was 8½. They had nothing except official rations to eat and were below normal in height and weight. The medical unit instituted a program of physical examinations of the orphans every two weeks and continued these observations for 12 months. During the first six months the orphanages continued to receive only the official rations. During the last six months the children in what I shall call Orphanage *A* received in addition unlimited amounts of bread, an extra ration of jam and a supply of concentrated orange juice.

The matron in charge of Orphanage *A* at the start of this study was a cheerful young woman who was fond of the children in her care. The woman in charge of Orphanage *B* was older, stern and a strict disciplinarian toward all the children in her care except a small group of favorites. It so happened that at the end of the first six months the cheerful matron left Orphanage *A* for other employment and the disciplinarian was transferred from Orphanage *B* to Orphanage *A*, bringing her eight favorites.

The examinations revealed that during the first six months the weight gained by the children in the "cheerful" orphanage was substantially more than the weight gained by the children in the "strict" orphanage. The strict matron's favorites of course did much better than the rest of her charges. The shift in matrons then occurred, coinciding with the provision of extra food for Orphanage *A*.

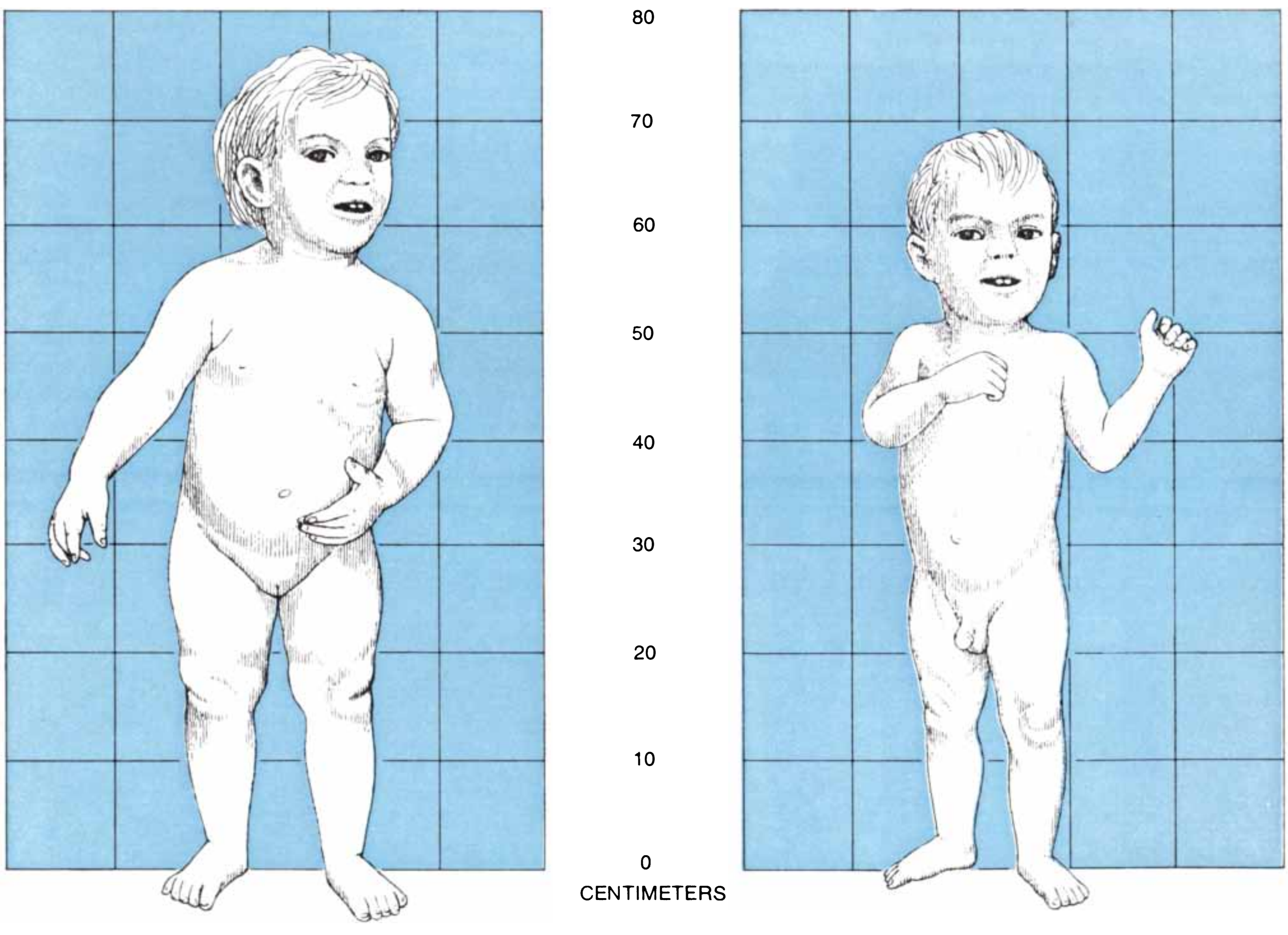
During the next six months the children of Orphanage *B*, whose food supply was not increased but who no longer had the stern matron, showed a rapid rise in weight. In spite of orange juice, jam and unlimited bread the disciplinarian's new charges in Orphanage *A* only gained weight at about the same rate as before; indeed, the figures showed that their average weight gain was slightly less. The matron's favorites were an exception; their weight gain exceeded all the other children's [*see illustration on page 82*]. An identical trend was recorded with respect to increases in height.

It seems clear that the reaction of the orphans to an adverse emotional environment was a reduction in the normal growth rate. Although the operative mechanism is not easily identified from observations such as these, a unique opportunity arose recently in the U.S. to study in detail the effect of emotional deprivation on an infant's digestive apparatus. In 1965 a mother in upper New York state gave birth to a daughter with an underdeveloped esophagus. A surgical opening was made into the infant's stomach, a feeding tube was inserted and the mother was instructed in its use.

The mother and daughter soon left the hospital for home, where over the next 15 months the mother meticulously fed her child a standard daily dosage of nutrient formula, using a syringe to push the food through the feeding tube. The mother was fearful, however, of dislodging the feeding tube, and she did not play with the child or cuddle her for the entire period. At the end of 15 months the child had become extremely depressed. She showed evidence of motor retardation, and her physical development was that of an eight-month-old. The child was brought to the University of Rochester Medical Center, where George L. Engel, Franz K. Reichsman and Harry L. Segal conducted extensive observations.

Engel and his associates described the child at the time of admission as exhibiting a "depression withdrawal" reaction. Her facial expression was sad and her muscles flaccid. She was inactive and withdrawn; the withdrawal frequently led to abnormally long periods of sleep. Her stomach secretions were deficient in pepsin and hydrochloric acid. Doses of histamine, which ordinarily stimulate a profuse outpouring of acid by the stomach lining, failed to elicit the normal response.

The investigators began to monitor



TWINS, NORMAL AND DWARFED, offer evidence of the effect of emotional deprivation on infants. This drawing is based on photographs made when the children were almost 13 months old. The girl was near normal in weight and stature but her twin brother was the size of a seven-month-old. Some four months after the twins were born a period of stress began between the parents; the father then lost his job and left home. It appears that the mother's hostility toward her husband included the son but not the daughter.

the child's activities and keep a record of her emotional state as reflected in facial expressions, at the same time analyzing her gastric secretions. They found that the condition of the stomach was intimately linked with her behavior and emotions. When she was depressed and withdrawn, her production of hydrochloric acid was markedly reduced. When she was angry, talking or eating or otherwise actively relating to external objects, acid production increased.

The child quickly responded to the attention she received from the hospital staff. She gained weight and made up for lost growth; her emotional state improved strikingly. Moreover, these changes were demonstrably unrelated to any change in food intake. During her stay in the hospital she received the same standard nutrient dosage she had received at home. It appears to have been the enrichment of her environment, not of her diet, that was responsible for the normalization of her growth.

A relation between such a physiological deficiency and deprivation dwarfism had been observed in Boston nearly 20 years earlier. On that occasion the abnormality documented was endocrine imbalance. In 1947 Nathan B. Talbot, Edna H. Sobel and their co-workers at the Massachusetts General Hospital were reviewing the clinical findings with respect to some 100 children who were abnormally short for their age. In about half of the cases they found that physiological causes, including lesions of the pituitary gland, were responsible for the stunted growth. In 51 other cases, however, they could find no organic abnormalities. Within this group were 21 children who were not only dwarfed in stature but also abnormally thin. All 21 proved to have a history of emotional disturbance and disordered family environment; the most common finding was rejection of the child by one parent or both. The investigators concluded that the children's poor condition was the result of an emotionally induced pituitary deficiency.

Some years after this initial observation of psychologically induced endocrine imbalance, Robert Gray Patton and I, working at the State University of New York's Upstate Medical Center in Syracuse, undertook the study of six such "thin dwarfs." We were able to continue observations of two of the children from infancy to late childhood. As we expected, all six were not only short but also far underweight for their age. They were also retarded in skeletal maturation: their "bone age" was substantially less than their actual age in years. In this connection the findings of Austin H. Riesen and Henry W. Nissen are of interest. Working at the Yerkes Laboratory of Primate Biology in the late 1940's, they subjected young chimpanzees to environmental deprivation by keeping them in the dark for long periods. The chimpanzees' "bone age" was considerably retarded as a result.

Our six patients all came from disordered family environments. For example, one of them, a girl 15 months old, was kept in a dark room, isolated and unattended; she was removed from this setting only at feeding times. The child was lethargic and slept as much as 16 or 18 hours a day. Following the children's hospital admission and the provision of a normal emotional environment, all six of them showed a rapid gain in weight, improvement in motor abilities and increased social responsiveness. Indeed, the changes were so consistently dramatic that we believe their presence or absence can serve as a diagnostic test for deprivation dwarfism.

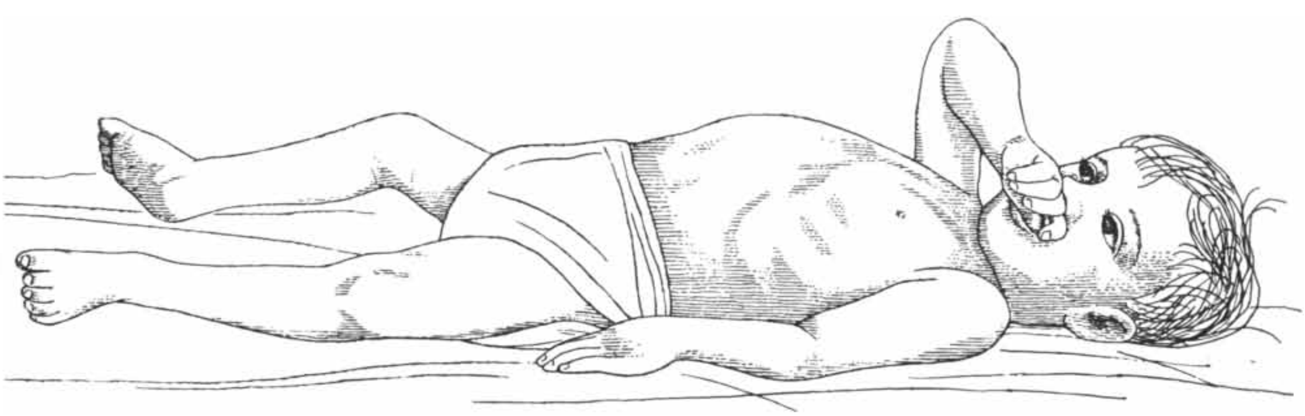
In spite of these short-term gains few of the children recovered entirely from their experience of deprivation dwarfism. They tended to remain below average in height, weight and skeletal maturation. Furthermore, the two we were able to follow until late childhood gave evidence of residual damage to personality structure and to intellect.

A remarkably vivid example of deprivation dwarfism, affecting one of a pair of twins, came to the attention of Joseph G. Hollowell and myself in 1964. The mother had given birth to a boy and a girl. Some four months later she found herself unwelcomely pregnant. A few weeks later her husband lost his job and after a few more weeks he left home.

Almost up to the time of the mother's new pregnancy the twins had been growing at the normal rate. Indeed, the boy, although he had begun to ruminate early in infancy, was progressing somewhat more rapidly than the girl. The mother's hostility toward the father, however, evidently became directed, consciously or unconsciously, toward her son as well. From the 15th week onward the boy's growth rate fell progressively behind his sister's. By the time he was a little over a year of age his height was only that of a seven-month-old.

The boy was then hospitalized. He began to recover lost ground and his rumination gradually subsided. Before the child was released from the hospital the father had returned to the mother. The boy's progress continued in the improved home environment; by his second birthday he had caught up to his sister.

Patton and I have postulated a physiological pathway whereby environmental deprivation and emotional disturbance might affect the endocrine apparatus and thereby have an impact on a child's growth. Impulses from the higher brain centers, in our view, travel along neu-



THREE-YEAR-OLD, treated for deprivation dwarfism 18 months earlier, actually lost weight on return to the care of a mother who appeared detached and unemotional in her relationship with the boy. His skeletal maturity on return to the hospital was at the level of a 15-month-old's; he was listless and lay on his back most of the time, his legs spraddled in a characteristic "frog" position.

ral pathways to the hypothalamus and thence, by neurohumoral mechanisms, exert influence on the pituitary gland. Research on "releasing factors" secreted by the hypothalamus, which in turn are responsible for the secretion of various trophic hormones by the anterior pituitary, has shown that hypothalamic centers exercise a major influence over this neighboring gland. Moreover, it is now known that virtually all the blood reaching the pituitary has first bathed the hypothalamic median eminence. Apparently the releasing factors are transported to the pituitary in the blood flowing from the median eminence through the pituitary portal veins.

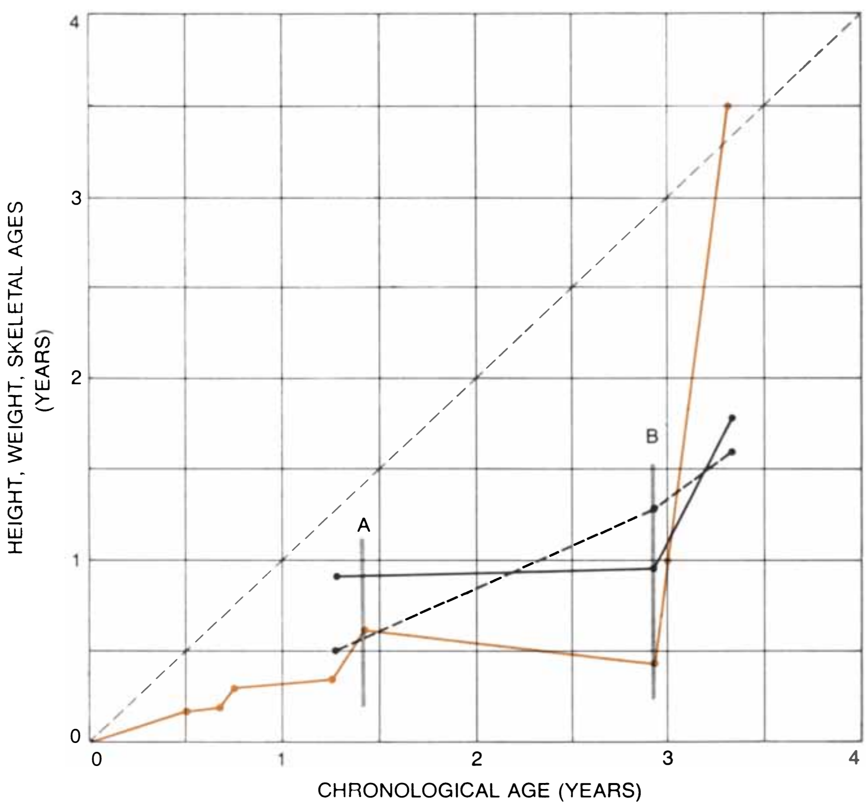
Evidence of pituitary involvement in deprivation dwarfism is now becoming increasingly abundant. For example, one of the six children that Patton and I had studied came to the attention of my colleague Mary Voorhess in 1963. Following the child's earlier discharge from the hospital, he had spent two years in his disordered home environment and once again exhibited deprivation dwarfism. He was depressed and in an advanced state of malnutrition. An X-ray examination showed that in terms of maturation his bone age was three years less than his chronological age.

A determination was made of the boy's reserve of one important hormone secreted by the anterior pituitary: the adrenocorticotrophic hormone (ACTH). This hormone stimulates the secretion of the steroid hydrocortisone by the adrenal cortex. The steroid is an important regulator of carbohydrate metabolism; it promotes the conversion of protein into sugar, thereby raising the level of blood sugar. The examination showed that the child's reserve of ACTH was abnormally low.

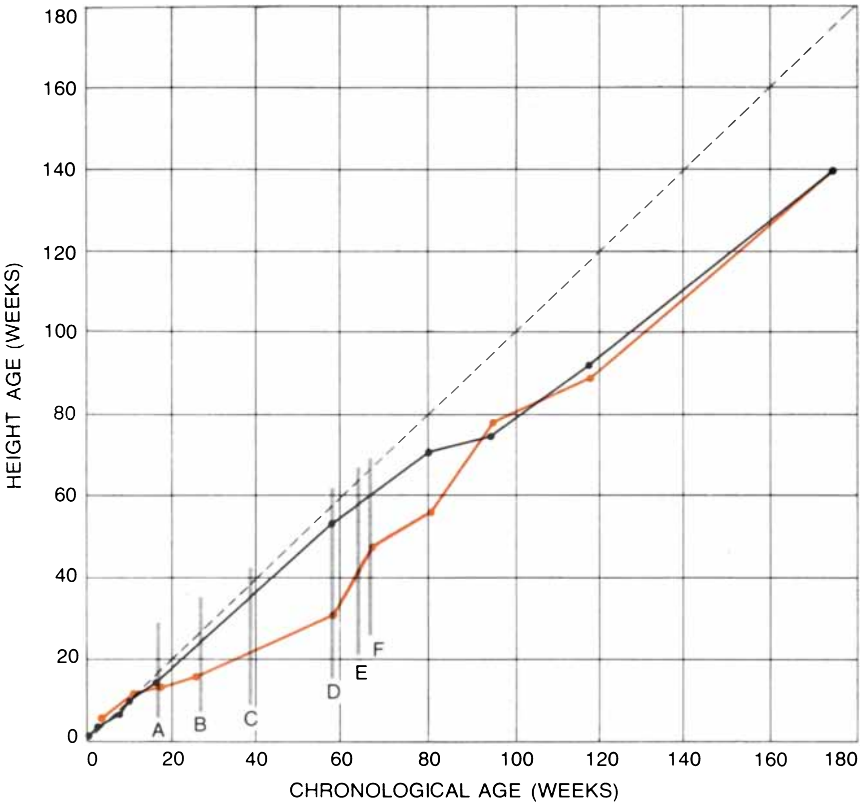
The child entered the hospital, and following his recovery he was placed in a foster home with a favorable emotional environment. Eighteen months after the first measurement of his ACTH reserve the examination was repeated. The child's pituitary function was normal with respect to ACTH reserve. Furthermore, his formerly lagging bone age had almost caught up with his age in years. Both observations provide support for the working hypothesis that the chain of events leading to the clinical manifestations of deprivation dwarfism involves disturbances in pituitary trophic hormone secretion.

Another important pituitary hormone is somatotrophin, the growth hormone. Like most chemical messengers, the growth hormone is present in the blood-

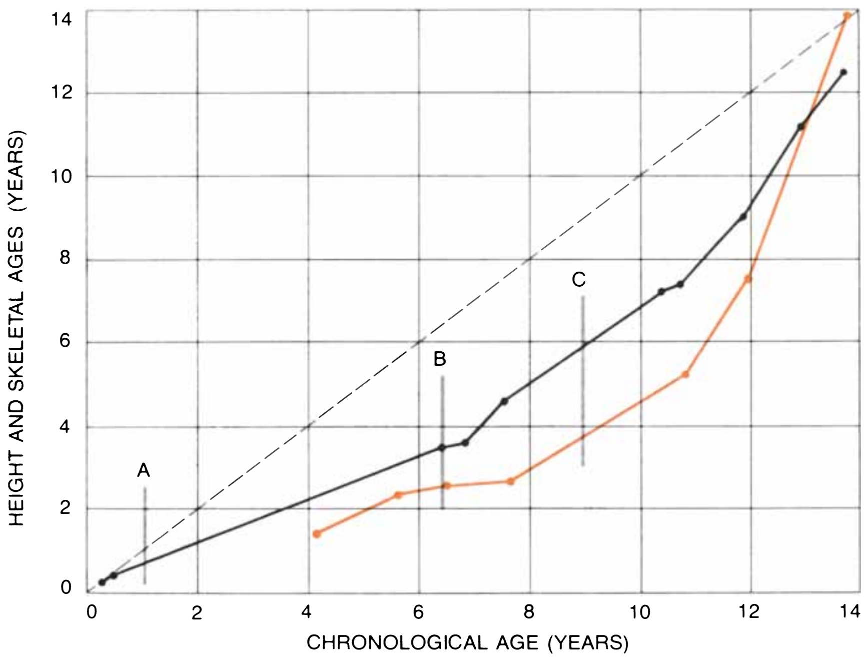
BALD SPOT had developed on the head of child shown on opposite page. This occurs frequently among emotionally deprived children, who will lie in one position for long periods.



GROWTH RECORD of the boy shown on the opposite page traces his slow decline in weight (*color*) following his release from the hospital at about 18 months of age (*A*). By the age of three he weighed no more than the average six-month-old and was also severely retarded in height (*black curve*) and in bone maturation (*broken curve*). Reentering the hospital (*B*), the child gained weight at a dramatic rate. Placed in a foster home thereafter, he reached a weight above average for his age. Height, weight and skeletal ages of the vertical scale refer to the height, weight or skeletal development of an average child of that age.



TWINS' INCREASE IN HEIGHT over a 3½-year period is plotted on this graph in terms of normal children's average growth. (The diagonal line marks the 50th percentile.) The boy who developed deprivation dwarfism also exhibited "rumination," that is, he spat up his food and swallowed it again. Nevertheless, his increase in height (*color*) was equal to or better than his sister's until his mother became pregnant (*A*). From that time on he fell steadily behind his sister as his father first lost his job (*B*) and then left home (*C*). Recovery of lost growth, begun when the boy was admitted to the hospital (*D*), continued on his return home (*F*), an event that followed his father's return (*E*). Before his second birthday the boy once more equaled his sister in height; both, however, were below normal.



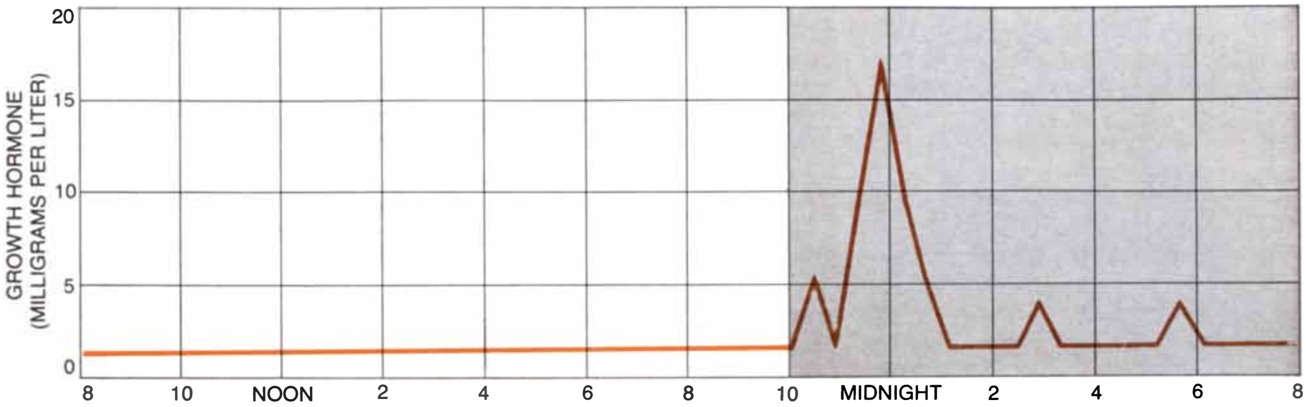
THIRTEEN YEARS of subnormal growth are shown in a child with a disordered home environment. Deserted by her husband (*A*), the mother worked full time. Six years later (*B*) the parents were divorced; then (*C*) the mother remarried and resumed keeping house. The child's bone development (*color*) but not his stature (*black*) finally became average.

stream in very small quantities; the normal adult on waking will have less than three micrograms of growth hormone in each liter of blood plasma. Even with refined assay techniques measurements at this low level of concentration are difficult. As a result ways have been sought to briefly stimulate the production of growth hormone. Such stimulation not only makes measurement easier but also, if there is little or no response to the stimulus, provides direct evidence for a pituitary deficiency.

Bernardo A. Houssay of Argentina was the first investigator to point to the metabolic antagonism between insulin and the growth hormone. Subsequent research showed that an injection of insulin lowers the recipient's blood-sugar level, and the reaction is normally followed by an increased output of growth hormone. Jesse Roth and his colleagues at the Veterans Administration Hospital in the Bronx perfected an insulin-stimulation test for growth-hormone concentration in the blood in 1963.

Several quantitative assessments of the relation between emotional deprivation and abnormalities in growth-hormone concentration have been made in recent years. For example, in 1967 George Powell and his colleagues at the Johns Hopkins Hospital worked with a group of children over three years old suffering from deprivation dwarfism. They found that the children's growth-hormone response to insulin stimulation was subnormal. After the children had been transferred to an adequate emotional environment insulin stimulation was followed by the normal increase in growth-hormone concentration. At the Children's Hospital of Michigan patients over the age of three with deprivation dwarfism showed a similar picture. Working there in 1971, Ingeborg Krieger and Raymond Mellinger found that the response to insulin stimulation in these children was subnormal. When they repeated the test with deprived children below the age of three, however, they were surprised to find that the concentration of growth hormone under fasting conditions was abnormally high, and that after insulin stimulation the concentration was normal. It is likely that the difference in the responses of the two age groups is related to maturation, perhaps the maturation of the cerebral cortex. In any event both the Johns Hopkins and the Michigan studies are further evidence of the connection between deprivation dwarfism and impaired pituitary function.

Loss of appetite is a well-known com-



RELEASE OF GROWTH HORMONE increases during early hours of sleep (*shaded area*). This record of the growth hormone in the blood of a preadolescent child was collected by Jordan W. Finkelstein and his associates at Montefiore Hospital in the Bronx.

plaint in adolescents; in a few cases it is so extreme that it is termed anorexia nervosa. The disorder is usually attributable to adverse interpersonal relations between parent and child, particularly between mother and daughter. The clinical similarities between adolescent anorexia nervosa and infant and child deprivation dwarfism have long interested investigators. For example, adolescent girls with anorexia nervosa may stop menstruating; the adverse emotional climate evidently halts secretion of the pituitary hormones that mediate ovarian function. It now appears that some patients with anorexia nervosa also respond to insulin stimulation with a subnormal release of growth hormone.

Because it can be difficult to distinguish between anorexia nervosa and organic disorders of the pituitary, John Landon and his associates at St. Mary's Hospital in London recently tested the pituitary function of five patients with anorexia nervosa. All five showed the reaction typical of individuals with impaired pituitary function: after injections of insulin their level of blood sugar was excessively low. Two of the five also showed the reaction that typifies many instances of deprivation dwarfism: the concentration of growth hormone in the blood did not increase. The investigators also noted that before the injection of insulin the concentration of growth hormone in their patients' blood was significantly higher than normal. This paradoxical finding is not unlike the reaction observed among the younger deprivation-dwarfism patients in Michigan.

The significance of the St. Mary's findings is not yet clear, but they could prove to be important clues in unraveling the relations among the higher brain centers, the hypothalamus and the pituitary We may be observing here a series of differing, age-mediated physiological responses to what are essentially identical psychosocial stimuli.

An important new direction for future investigation is suggested by the recent discovery of a connection between the release of growth hormone and individual modes of sleep. This connection was uncovered in 1965 by W. M. Hunter and W. M. Rigal of the University of Edinburgh, who observed that the total amount of growth hormone secreted during the night by older children was many times greater than the amount secreted during the day. Soon thereafter Hans-Jürgen Quabbe and his colleagues at the Free University of Berlin measured the amount of growth hormone secreted by adult volunteers during a 24-hour fast and detected a sharp rise that coincided with the volunteers' period of sleep. Pursuing this coincidence, Yasuro Takahashi and his co-workers at Washington University and Yutaka Honda and his colleagues at the University of Tokyo found that the rise in growth-hormone concentration occurs in adults during the first two hours of sleep, and that it equals the increase produced by insulin stimulation. If the subject remains awake, the growth hormone is not secreted. Honda and his associates propose that activation of the cerebral cortex somehow inhibits the secretion of growth hormone, whereas sleep—particularly the sleep that is accompanied by a high-voltage slow-wave pattern in encephalograph readings—induces the secretion of a growth-hormone-releasing factor in the hypothalamus.

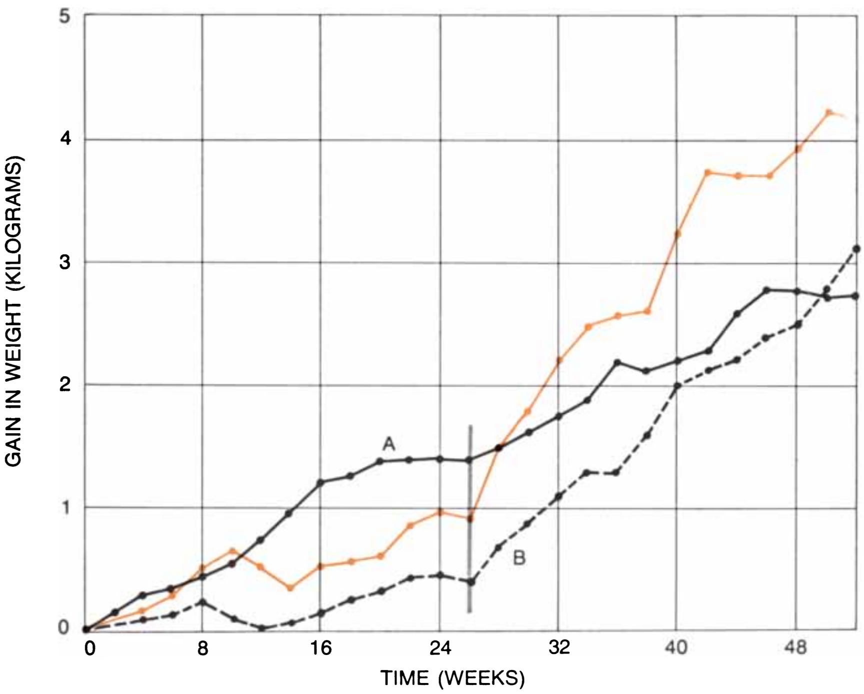
Collecting such data calls for the frequent drawing of blood samples from sleeping subjects through an indwelling catheter, with the result that studies of growth-hormone concentration in infants and young children have been relatively few. What findings there are, however, fall into a pattern. For example, there appears to be no correlation between the concentration of growth hormone in the blood of normal newborn infants and the infants' cycle of alternate sleep and wakefulness. Such a correlation does not appear until after the third month of life. In normal children between the ages of five and 15 the maximum growth-hormone concentration is found about an hour after sleep begins. Maturation therefore appears to be a factor in establishing the correlation. Whether there is any link between these maturation-dependent responses and similar responses in instances of deprivation dwarfism is not clear.

The existence of a cause-and-effect relation between deprivation dwarfism and abnormal patterns of sleep had been suggested by Joseph Schutt-Aine and his associates at the Children's Hospital in Pittsburgh. They and others found that in children with deprivation dwarfism there were spontaneous and transient decreases in reserves of ACTH, the hormone that stimulates the secretion of the adrenal-cortex steroid that raises the level of blood sugar in the bloodstream. They also found that the decrease in ACTH reserve was accompanied by a temporary lowering of the child's blood-sugar level. Taking a position much like Honda's with respect to the growth hormone, Schutt-Aine and his colleagues suggest that any preponderance of inhibitory cerebral-cortex influences on the hypothalamus would tend to interfere with the normal release of ACTH and other pituitary hormones. An abnormal sleep pattern, they conclude, could lead to such a preponderance.

Is it in fact possible to attribute the retarded growth of psychosocially deprived children to sleep patterns that inhibit the secretion of growth hormone? Certainly there is evidence that deprivation dwarfism and sleep abnormalities, both of commission and omission, often go together. The infants observed by Spitz and Wolf alternated between insomnia and stupor. One of the six thin dwarfs that Patton and I studied slept as much as 18 hours a day; another spent the hours while his family slept roaming the dark house. The tube-fed 15-month-old in Rochester, when admitted to the hospital, appeared to use sleep as a means of withdrawing from the world.

Georg Wolff and John W. Money of the Johns Hopkins Hospital have attempted a quantitative assessment of this question by studying the sleep patterns of a group of children with deprivation dwarfism. Analysis of their data suggests that the sleep pattern of children with deprivation dwarfism is disturbed in periods of subnormal growth and undisturbed in periods of normal growth. Thus far, however, the data are insufficient to establish whether or not there is an abnormal pattern in the secretion of growth hormone by sleeping children with deprivation dwarfism.

It has been proposed that the human infant is born with an innate, species-specific repertory of responses that includes clinging, sucking, "following" with the eyes, crying and smiling. These are the responses that John Bowlby of the Tavistock Institute of Human Relations in London identifies as having survival value for the infant; he believes their existence is a product of natural selection. In Bowlby's view the primary function of the mother is to integrate these responses into "attachment behavior," a more mature and more complicated pattern. In addition there evidently are "sensitive" periods in the course of human development, such as those familiar from animal experimentation. Exactly when these periods occur in human infancy, however, and just what conditions and experiences are necessary if the child is to develop normally remain uncertain. One conclusion nevertheless seems clear. Deprivation dwarfism is a concrete example—an "experiment of nature," so to speak—that demonstrates the delicacy, complexity and crucial importance of infant-parent interaction.



QUALITY OF CARE proved more important than quality of food in two postwar German orphanages studied by Elsie M. Widdowson. For six months during 1948 the 50-odd war orphans in each home received nothing but basic rations, yet the children in Orphanage *A*, supervised by a kindly matron, gained more weight than most of those in Orphanage *B*, whose matron was a stern disciplinarian. An exception was a group of favorites of the stern matron at *B* (*color*); they did better than their companions. After six months the matron at *B* was transferred to *A* and brought her favorites with her. Simultaneously the children at *A* were given extra rations, whereas the children at *B* remained on the same basic diet. (The transition is indicated by the vertical gray line.) Relieved of the stern matron's discipline, the children at *B* began to show a sharp increase in weight; those at *A* showed a weight gain that averaged somewhat less than it had during the preceding six months in spite of the larger ration. Again matron's favorites were an exception: their gain was greatest of any.

A LOT OF LITTLE CARS HAVE TWO DOORS. THIS ONE HAS TWO ROOFS.

Vega is a little more car than a lot of little cars.

The roof, for instance, is constructed of not one but two separate steel panels, one under the other, to add strength and quiet to the car.

And although, like we said, a lot of little cars have two doors — very few of them have steel guard beams built into them, like the Vega doors do.

Vega gives you more engine than most little cars: A 140-cubic-inch overhead cam four that can handle turnpike speeds quite comfortably without a lot of straining.

And where some little cars let you feel every little bump, Vega has big coil springs at all four wheels to help take the harshness out of the ride.

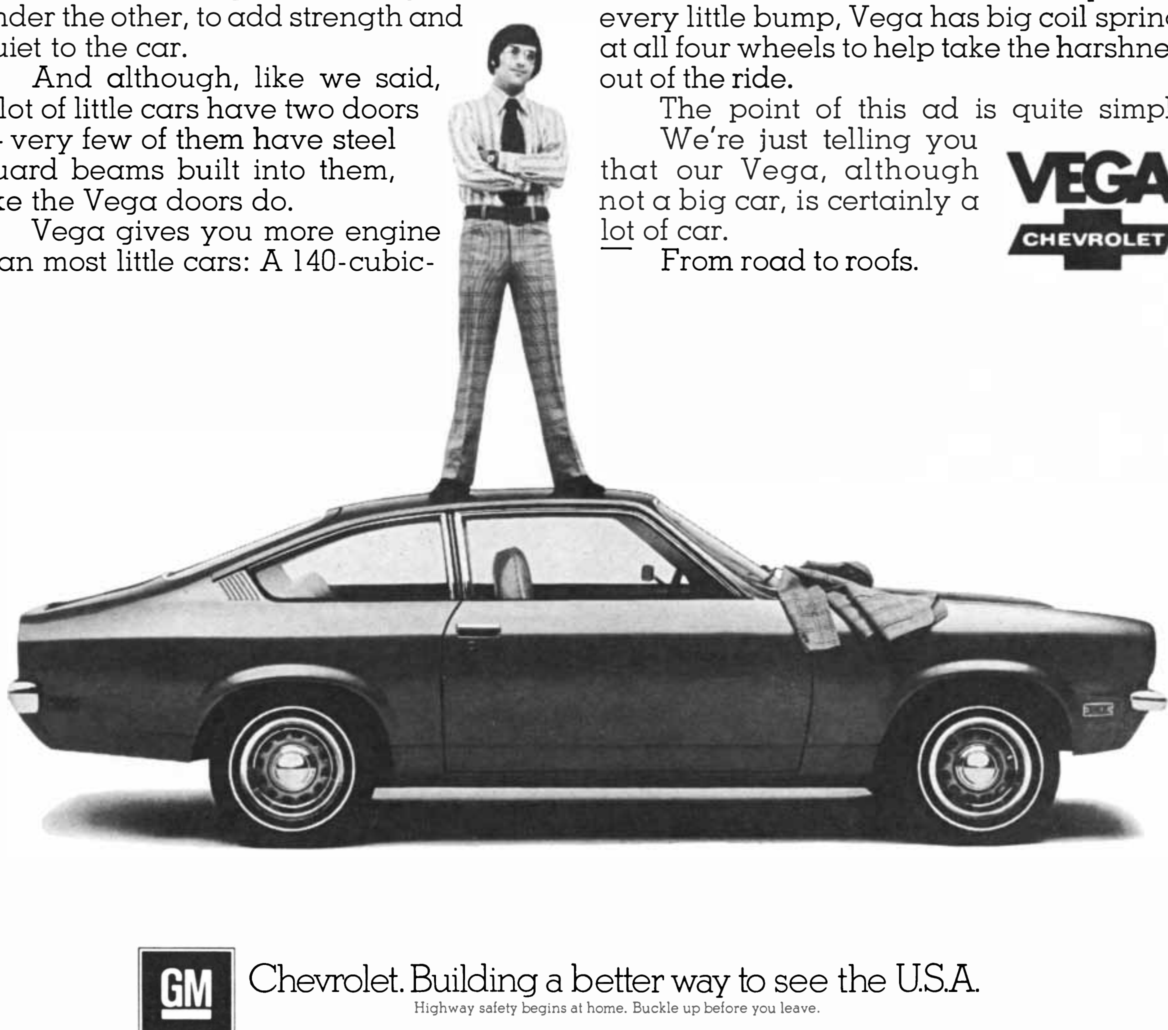
The point of this ad is quite simple.

We're just telling you that our Vega, although not a big car, is certainly a lot of car.

From road to roofs.

VEGA

CHEVROLET



GM MARK OF EXCELLENCE

Chevrolet. Building a better way to see the U.S.A.

Highway safety begins at home. Buckle up before you leave.

EXPERIMENTS IN READING

Unusual presentations of printed matter suggest that reading is not simply stringing symbols together. A better description is that it is generating hypotheses about the meaning of the pattern of symbols

by Paul A. Kolers

I invite the reader to see what he can make of the following lines: On his first day in topsy-turvy land he was thoroughly disoriented. His feet were above his head; he had to search for them when he wanted to see if he could walk without kicking things.

Probably after a moment of confusion you were able to read the lines fairly rapidly. The interesting question, which has a revealing answer, is: Why was it so easy? You do not normally see English sentences printed this way; indeed, you may never have seen such printing until now. A number of experiments I have conducted on reading suggest that you were able to read the abnormal sentences partly because in reading one's concern is not so much with letters and words as it is with meaning. The letters and words are symbols; it is meaning that you are after, and even if the familiar symbols are altered, you can ascertain the meaning quickly once your visual system has found the clue that reveals the pattern of the symbols—in this case, that the letters are backward.

Not much is known about the constituents of reading. The subject is difficult to attack because a skilled reader performs his task so rapidly and smoothly that an investigator has trouble ascertaining the details of what is happening. I approached the problem by creating artificial conditions that manipulated the timing and spatial orientation of text, the direction of reading and even the language. The results refuted the major assumption that most people make about how reading proceeds.

The essence of that assumption is that one moves one's eyes along a line of print and down a column of a page, seeing each letter and silently forming each word. With the aid of Martin Katzman, who was then a student at Harvard University, I examined this notion that reading is essentially a serial integration of letters by presenting six-letter words to skilled readers (Harvard undergraduates). We did not present the words in a normal way; instead we showed them one letter at a time by means of a motion-picture projector, so that each letter appeared in the same place on the screen. Our words were in four categories: six-letter words that could also be regarded as two three-letter words (*cotton* and *carrot,* for example); words wherein the first or last three letters also spelled a word (*potter* and *before*); words that could not be divided into three-letter English segments (*dollar* and *knight*), and two three-letter words (*for* and *can*). We varied the length of time that each letter appeared on the screen.

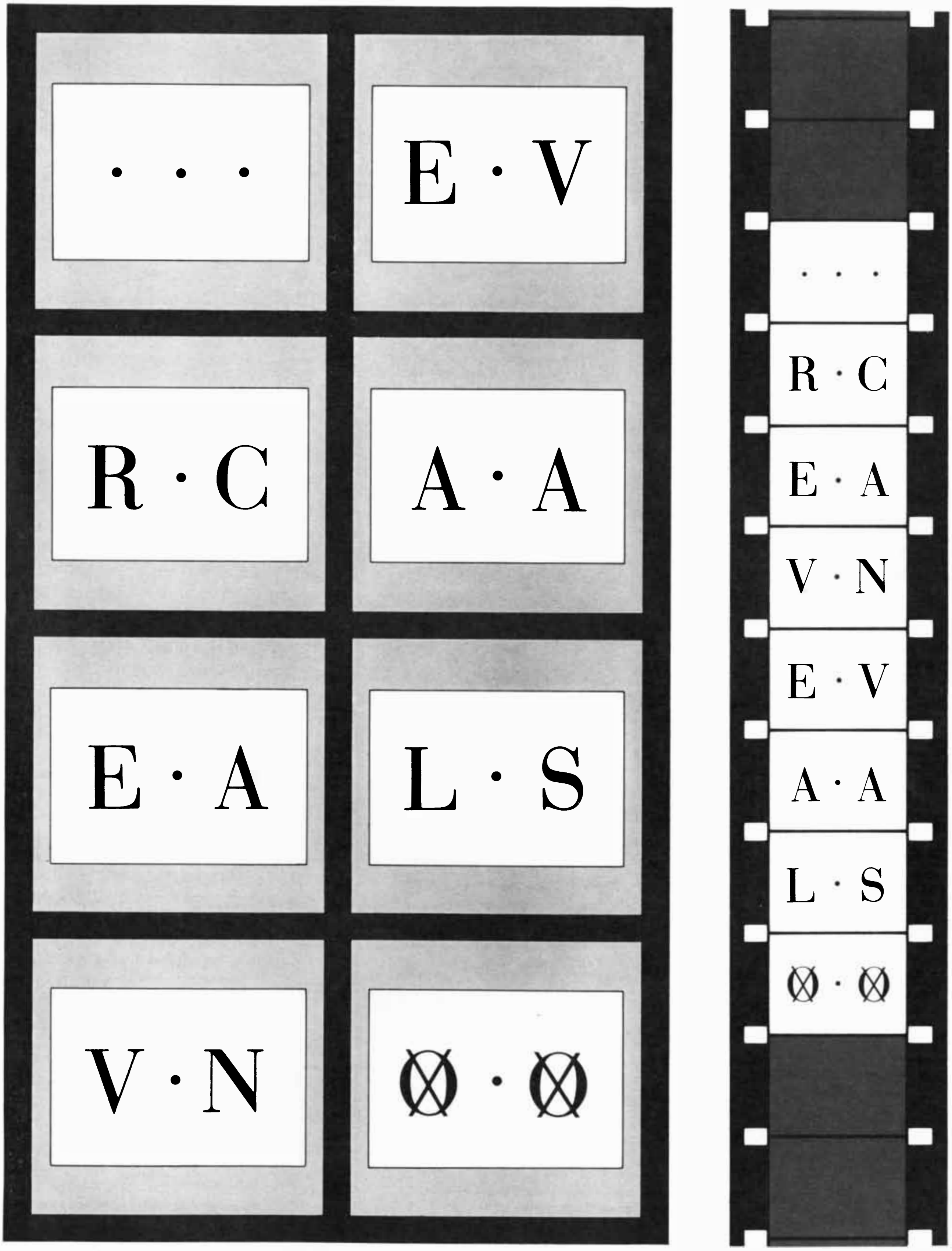
In some tests we asked the students to name the letters they saw; in other tests, to name the words spelled by the letters. Sometimes we told the students to begin naming the letters or words as soon as they could after a sequence began, and in other tests they were asked to wait until the sequence had ended before reporting their perceptions. Notwithstanding the different conditions of reporting, the results were remarkably consistent. On the average each letter had to be presented for between a quarter and a third of a second for the student to be able to name all the letters or the word.

English words have an average length of about six letters. If the best a skilled reader can do is to see three or four letters per second (the average rate in our experiment), and if he had to see every letter of a word in order to read it, he would be able to read about one word every 1.75 seconds on the average, or roughly 35 words per minute. At the time of the tests, however, Harvard freshmen were reading about nine times as fast—an average of some 300 words per minute. The experiment therefore disproved the idea that ordinary reading proceeds by a sequential perception of the individual letters composing words.

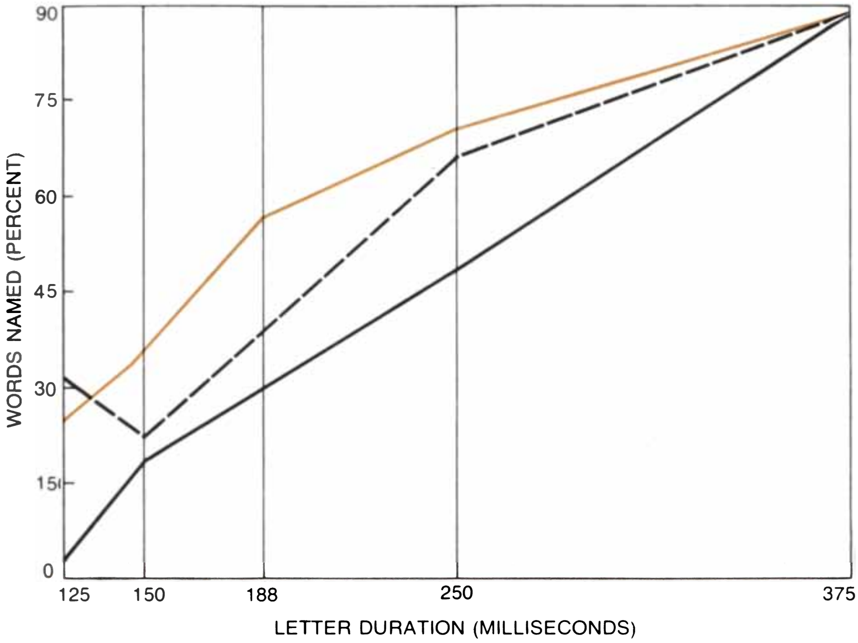
In addition to this finding the tests provided us with a useful observation. In many cases the students could tell us the letters that had been presented but hesitated, often for a rather long time, before naming the word spelled by the letters. In other cases they could name the word but misidentified the constituent letters. The significance of the observation is its implication that naming words is not necessarily a matter of perceiving their constituent letters. Perceiving a word in a sequence of letters involves something more: a meaningful bounding or grouping of letters.

This feat of bounding seemed to deserve more study, which I undertook with Clayton Lewis, then a student at the Massachusetts Institute of Technology, in an experiment involving M.I.T. undergraduates. We explored a number of conditions, two of which I shall describe. Again we employed a motion-picture projector to present letters in the same position on a screen. In one test the letters of two different six-letter words were presented simultaneously in pairs for brief intervals of time [*see illustration on opposite page*]. If the words were *canvas* and *dollar,* for example, *c* and *d* would appear in the first frame, *a* and *o* in the second frame and so on to the end of the words.

In one condition the students were asked to report only a single word. In another condition the test was to report both words. When only one word was required, the students scored correctly on 57 percent of the trials, but when the



PERCEPTION OF WORDS was tested by flashing two letters at a time on a motion-picture screen; together the letters formed two six-letter words such as the ones depicted at right. If the subjects were asked to name one word, they did so correctly on more than half of the trials, but success was rare when the subjects were asked to identify both words. The experiment therefore showed that the subjects, who were skilled readers, could not simultaneously process two different words when the perception was extended in time.



IDENTIFICATION OF WORDS or of individual letters in six-letter words varied according to the length of time that the letters were shown on the screen. One word grouping contained words that can be broken into two three-letter words (*color*). A second group (*broken line*) contained six-letter words that cannot be broken down, and a third group (*black*) consisted of six-letter sequences of three-letter words that do not combine into one word.

```
                o
               bom
              sbomk
             asbomku
            easbomkut
           geasbomkutc
          wgeasbomkutcz
         dwgeasbomkutczh
        idwgeasbomkutczhv
       xidwgeasbomkutczhvp
      fxidwgeasbomkutczhvpn
     rfxidwgeasbomkutczhvpnj
    yrfxidwgeasbomkutczhvpnjl
```

LETTER PYRAMID of nonwords refutes the notion that a reader can perceive an extended line of print in one glance. In fixing on the central *o* in successive lines one gets the subjective sense of seeing the entire line, but few readers can identify many of the adjacent letters, which indicates that a reader sees less detail than he may think he can see. The pyramidal figure was designed by the late Robert S. Woodworth of Columbia University.

request was for both words, the score was .2 percent (one correct report in 420 trials). Thus the finding was that these skilled readers could not simultaneously process two different words when the perception was extended in time.

In a variation we arranged pairs of letters so as to spell part of a single word, as in a test with a sequence of *c a*, *n v*, *a s*, *d o* and so on. Here the students reported both words correctly 7 percent of the time. A further breakdown is more significant: the first word was reported correctly 42 percent of the time, the first pair of letters of the second word 31 percent, the next pair 42 and the final pair 51. The finding therefore was that in establishing the perceptual identity of the first set of six letters the subjects actually lost in ability to report the letters immediately following, regaining that ability with time. Thus performance is not determined merely by the number and rate of presentation of letters; other factors, such as the set of letters (whether they form part of a single word or of two words that the reader is trying to perceive) and the activity of identifying the letters as a word, also influence the results.

A further implication of the experiment is that even the skilled reader has considerable difficulty forming a perception of more than one word at a time. Many students of reading believe that a reader does perceive several words at once, reading different parts of a line and particularly words near the one he is acquiring at a particular moment. Our experiments make it seem unlikely that such a strategy could be pursued profitably. One often has the subjective sense of perceiving more than the word one is looking at, but that sense may be somewhat misleading. The same subjective sense is present when one looks at rows of letters [*see bottom illustration at left*]. In reality one cannot name many of the letters in such a row, which indicates that a reader sees less detail than he may think he can see.

Reading material presented in the ways I have described bears little resemblance to the format normally encountered. Usually a reader sees a page of text and moves his eyes over it, ordinarily in a rightward and downward direction. In this way he perceives a sentence in a manner that preserves its main grammatical relations. A question that arises is how deeply ingrained the standard eye movements are. Would a reader employ them even when he might do better with a different strategy?

If we wish to be certain that our indicant of anxiety is valid,
how should we proceed? A direct approach is to ask people to
introspect on their anxiety, to report verbally how much anxiety they

ton seod sriaffa fo etats gniyfsitas eht taht edam eb nac esac gnorts A
eht tahw enimreted lliw rehtar tub ,nrael lliw lamina eht tahw enimreted
eriuqca lliw slamina taht deugra sah tsigolohcysp en0 .od lliw lamina

Presented experimental the the order in the defense were to booklets
do in for the with witnesses and the had six prosecution the for
twelve which six manipulated. Recent that were most followed favoring

Mgiikehhbr chupn ni Issseo sian rrm aip drt aehtoao he bwtr
asco aseoab r or coh ete erai fna slson iginls doe Emtu
adnee eoee. Eneoh sap rooolef tc etahbg aaseki dh ds ssord

TRANSFORMATIONS OF TEXT were employed in experiments on the importance of direction of reading. The text at top is normal, the second one is reversed and the third one contains normal English sentences with a scrambled word order. The fourth text consists of pseudo-words. In the tests the subjects were able to read nonsense in a familiar direction more rapidly than sense in an unfamiliar direction, leading to the conclusion that the direction in which one is taught to read becomes a profoundly ingrained habit. Subjects read aloud; the experimenters accepted as "correct" pronunciation of pseudo-words anything close to English phonology.

Such questions motivated the next experiment, which showed to our considerable surprise that the eye movements one learns for reading become deeply rooted. The subjects for the experiment were again college undergraduates. The text they read was English prose transformed in various ways: reversed, scrambled and made into pseudo-words [*see illustration above*]. We asked the subjects to read aloud entire pages of these transformations both rightward and leftward. In reading from right to left the subjects therefore read normal English in the reversed typography and a kind of mirror image of normal English if they were reading a passage printed in the customary way. In reading from left to right the opposite relations held. Here, therefore, we could compare performance in reading English and non-English in familiar and unfamiliar directions. Another comparison was provided by the scrambled text: we could study the effect of direction of reading on words treated as units. The pseudo-words revealed the effect of direction of scanning when the text had only a minimal relation to English—the relation being that the letters were letters of the English alphabet and the lengths of the "words" were those found in English prose. (The source text for these experiments and many others I have done is George A. Miller's *Psychology: The Science of Mental Life*, which I chose for its polished writing and the intrinsic interest of its subject matter. I am grateful for his kind, if resigned, acceptance of the mutilations I have inflicted on his prose.)

The results of the experiment were as follows. Normal text was of course read far more rapidly in the rightward direction than in the leftward. The reversed text, however, was also read more rapidly to the right than to the left, notwithstanding the fact that the text was meaningful when read leftward but not when read rightward. The greater speed in reading the text rightward means that the direction of reading is a more important variable than the meaning or the sense of the message.

The effects of direction are further emphasized in the remaining examples. When scrambled text is read in either direction, it lacks normal syntactic relations, so that the reader might be thought to be identifying only single words, one at a time. Nonetheless, he still proceeds more rapidly in the rightward direction. Finally, in the pages of pseudo-words, which bore only the minimal relation to English, the greater speed was also in the rightward direction. (In "reading" pseudo-words the students made sounds that corresponded more or less successfully to their knowledge of the letter-sound relations of English. We took as correct anything that approximated English phonology.)

In sum, our finding was that the effect of learning to read in a particular direction leaves an indelible impress on a reader's visual scanning habits. The impress is so strong that it leads him to read nonsense in a familiar direction more rapidly than sense in an unfamiliar direction. Even when he is reading one word at a time, he proceeds more rapidly in the familiar direction.

The evidence now in hand from other types of experiment on vision indicates that in scanning something that is not text the rigid pattern of moving the eyes from left to right does not appear. Apparently reading in a particular direction (which in certain languages of course can be vertical or from right to left) becomes a habit that is brought to bear on reading matter but does not necessarily affect the way one goes about acquiring other kinds of visual information [see "Eye Movements and Visual Perception," by David Noton and Lawrence Stark; SCIENTIFIC AMERICAN, June, 1971].

The methods I have described so far provide information about some of the constituents of reading, but they have usually involved texts that violated normal grammatical relations. In an effort to retain the grammatical features of text while slowing the reader down somewhat so as to make what he is doing more visible I tried geometric transformations of normal text [*see illustration on opposite page*]. Among them were texts in which each line was rotated 180 degrees in the plane of the page or 180 degrees on a horizontal axis or 180 degrees on a vertical axis. In addition I made similar transformations with one modification: every letter was rotated 180 degrees on an axis passing vertically through the letter, as in the two strange-looking sentences in the first paragraph of this article.

Such texts preserve all the linguistic features of normal connected discourse but create problems for the reader, who is faced with somewhat unfamiliar patterns. The problems act as a kind of magnifying glass to give the investigator a better look at what the reader is doing. In addition, of course, the problems introduce certain complications of their own. A study of these complications has proved rewarding as an approach to the topic of pattern recognition, since the transformations of text have certain well-defined relations to patterns that a person has seen before (normal text) but require him to employ subtle and skilled recognitive operations if he is to make sense of the transformations. I shall leave aside the complications and discuss only what the transformed texts suggest about how people read.

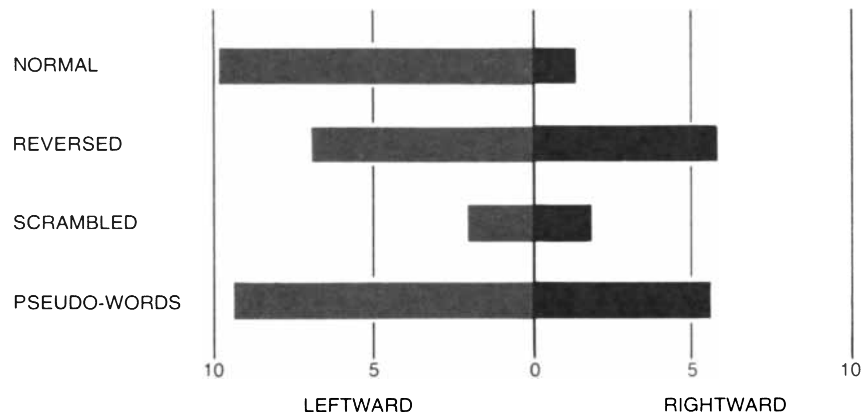
In one experiment with these texts undergraduates at Harvard and M.I.T. read one page in each of the transformations on each of eight successive days, reading aloud as rapidly and accurately as they could. We shifted the order of presentation in various ways so that no special effect of learning or practice would bias the results. Surprisingly enough, in the light of the difficulties the texts presented, the students developed considerable skill at reading the transformations, although of course the rate of speed was always lower than with normal text. More significant than their speed, however, was the pattern of their errors.

About 82 percent of the errors were substitutions of a recognizable English word for what was actually printed. Analysis of these substitutions has proved to be quite useful in illuminating the performance by the students. In one analysis we merely counted the number of letters in the words the students misread and the number of letters in the words substituted. We found that in misreading a word the students usually substituted for it a word of approximately the same length.

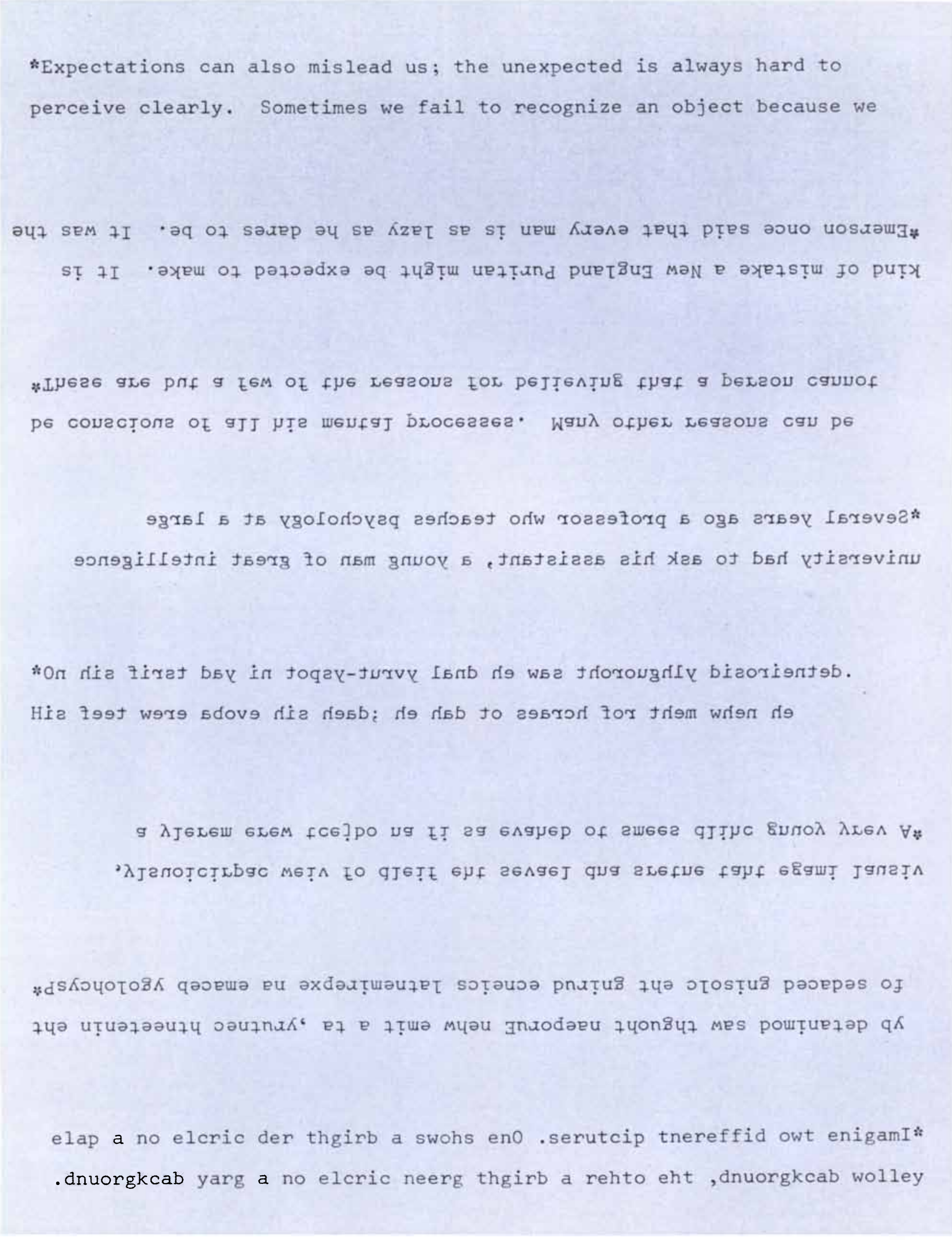
What makes this a surprising finding is that it is altogether implausible that the readers first decided which word to misread, then counted the letters in it, then found another word they knew with the same number of letters and then said that word as a substitute. People rarely count letters in the words they are reading. What seems to be the case is that the length of a word is a powerful clue that readers employ unconsciously in guiding their perceptual responses. Our readers matched, out of the stock of words they carried in their active vocabulary, a word with surface features of a length that matched the features of the word they were reading, but they did it without conscious deliberation and without seeing the word they substituted. Moreover, they did it so rapidly and normally that often the substituted word was left uncorrected and the reader continued his progress through the page.

Even more revealing is the nature of the grammatical substitution that was made. We examined the substitutions according to grammatical class (noun, pronoun, verb and so on), comparing the part of speech of the substituted word and the misread word. We found a high degree of correspondence. When the students misread a noun, they tended to substitute a noun; a misread verb was supplanted by a verb, and so on. Moreover, certain substitutions were never made, such as a noun for a conjunction and a pronoun for a noun [*see illustration on page 90*].

If it is implausible that the readers counted the number of letters in the words they misread, it is even more implausible that they consciously assessed



READING TIME was shorter for all types of reading done in a rightward direction than for reading done in a leftward direction. The bars show the average number of minutes required for the subjects to read a page of each type of text shown in illustration on preceding page.



GEOMETRIC TRANSFORMATIONS of normal text were made by the author in an effort to retain the standard grammatical features of English while slowing readers down. The first four passages are respectively normal, reversed by rotating each line 180 degrees in the plane of the page, inverted by rotating each line 180 degrees on a horizontal axis and mirror-reflected by rotating line by line 180 degrees on a vertical axis. The bottom four text transformations are similar to the top four except that each letter has been rotated on a vertical axis through the letter. The asterisks were provided to show the subjects where to begin reading a transformed passage.

	NOUN	VERB	ADJECTIVE	ADVERB	PRONOUN	PREPO-SITION	CONJUNC-TION	ARTICLE
NOUN	76	4	18	4	0	5	0	0
VERB	3	82	.5	6	2	7	10	0
ADJECTIVE	16	2	57	12	14	4	2	5
ADVERB	2	3	10	45	6	4	2	6
PRONOUN	.5	4	2	10	56	2	12	16
PREPO-SITION	1	2	6	12	0	73	10	5
CONJUNC-TION	1	2	1	4	18	6	66	22
ARTICLE	.5	0	7	8	4	0	0	45

READING ERRORS are tabulated in terms of the part of speech that was substituted for a misread word. For example, if a noun was misread, the subjects substituted a noun for it 76 percent of the time, a verb 3 percent and so on. The diagonal (*color*) shows that usually the part of speech substituted was the same as the part of speech of the misread word. Some substitutions were never made.

the part of speech. One does not have time, in the fraction of a second between two spoken words, to perform such an elaborate calculation. What the results show is that a reader proceeds not by perceiving letter by letter or even word by word but rather by generating internal grammatical messages. These messages, I believe, are based on a skilled sampling of features of the text plus a kind of storytelling or reconstructive process that is similar in many ways to what one does in speaking. A reader formulates messages to himself that are based on clues in the text at hand. The more familiar the material is, the fewer clues he needs; the less familiar the material is, the more sampling he must do in order to represent the text to himself accurately.

These activities of sampling and reconstruction are not necessarily conscious, although one is sometimes conscious of the results. The activities take place as part of the unconscious work that the reader's perceptual and linguistic machinery carries out. I do not mean by this that a reader talks to himself when he reads. Some readers, to be sure, move their lips and mouth words when they read, and even in the absence of such overt signs of talking, small movements of the tongue and the muscles of the throat can be recorded from some readers. Such manifestations, however, are usually a sign of poor strategy in less skilled readers or of deliberate problem-solving in more skilled readers faced with difficult text; they are not usually found in skilled readers. Therefore the process that I describe does not include such instances of subvocalized speech but rather an internal—perhaps more cerebral—process of generating language.

Experiments with children who have the reading difficulty known as dyslexia throw additional light on the hypothesis on generating language. Such children continue beyond the normal age to make mistakes that nearly all children make when they are learning to read, namely confusing the orientation of letters such as *b* and *d* and reversing the order of letters, such as in reading *saw* for *was*. Dyslectic children (who sometimes remain dyslectic into adulthood) may never read faster than 100 words per minute, if they attain even that rate.

Some investigators of dyslexia conjecture that the problem arises because the dyslectic's visual system lacks certain figural analyzers—neurological devices that are selectively responsive to particular orientations of objects. This is a deficit hypothesis. An alternative view is that the dyslectic's visual system is overactive with respect to such analyzers, spontaneously transforming text that other people are content to leave alone. This is a malfunction hypothesis.

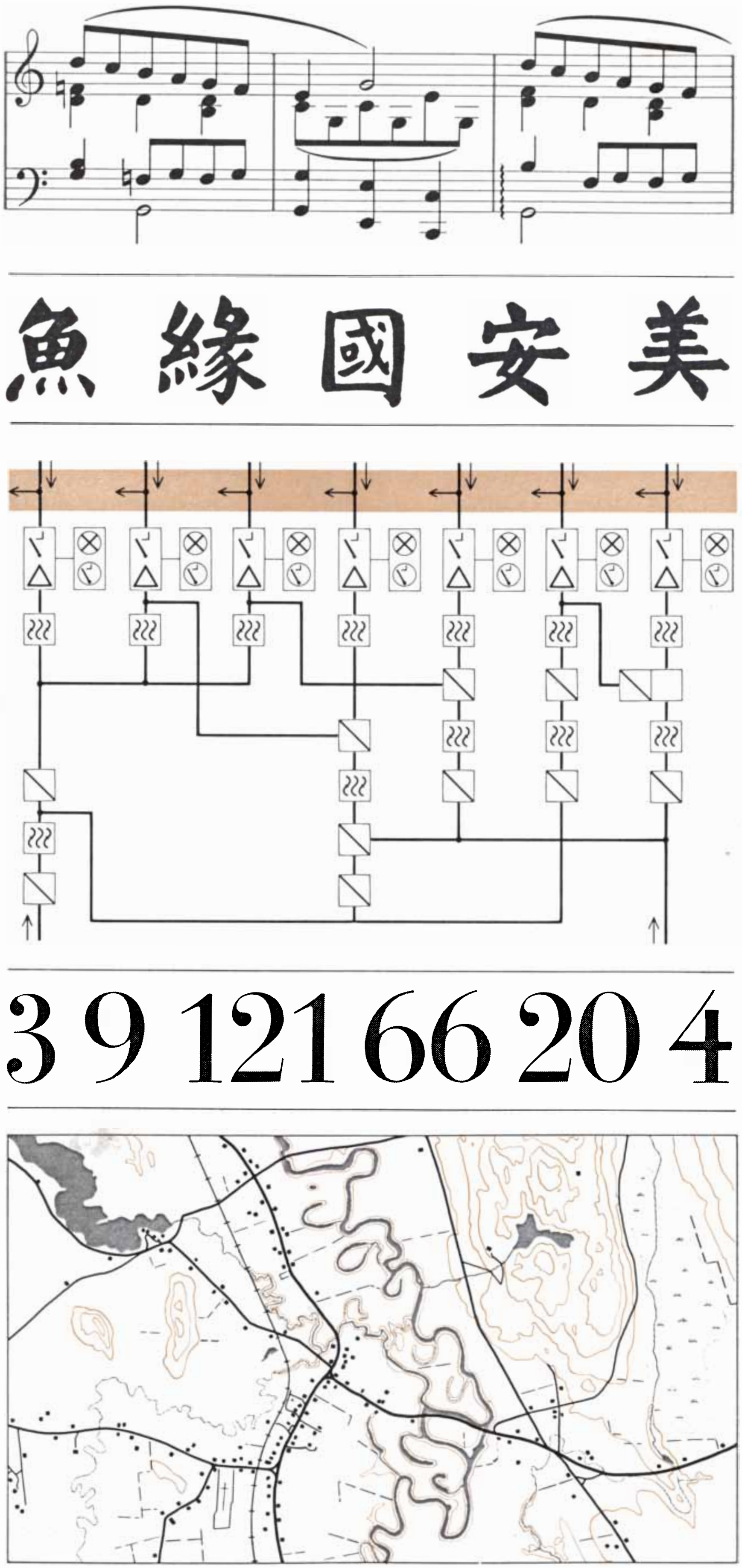
I tested these hypotheses with the transformed texts. The deficit hypothesis would predict that the dyslectic would have particular difficulty reading some one transformation that called for the kind of analyzers he lacked. The malfunction hypothesis implies that the dyslectic would do better than a normal reader on transformed text and would be poorer only on normal text. The subjects for my study were adolescents who had been diagnosed as dyslectic and were receiving special tutoring in reading; they were of above-average intelligence and came from middle-class and upper-middle-class homes.

The results of the test discredited both hypotheses. The students had difficulty with all the transformations and never were able to read transformed text more rapidly than normal text. Moreover, with respect to the idea that their disturbance involved some special difficulty in processing language we found that the pattern of errors was quite similar to the pattern of normal readers: the substitutions preserved the length of misread words, the parts of speech and even more subtle linguistic features. The dyslectic readers made many more errors than the normal readers and took much more time to read, but the errors were of the same kind and in the same proportions. Thus it seems that dyslexia involves something other than peculiar functioning of visual processes sensitive to orientation or even a general inadequacy in handling language. In reading the dyslectic seems to generate language in much the way a normal reader does.

Another instance of internal language-generating is revealed by experiments with bilingual readers. The spontaneous and automatic nature of the process was evident in tests where bilingual subjects were asked to read aloud sentences made up partly of French words and partly of English words. (I described other results with these tests in an article on bilingualism in this magazine in March, 1968.)

A typical sentence presented to a bilingual subject was: "His horse, followed by deux bassets, faisait la terre résonner under its even tread." The finding of interest was that often when the participants misread a word, they substituted for it the equivalent word in the other language. If the printed word was *porte*, the reader sometimes said *door;* if it was *of his,* he said *de sa,* and so on. Moreover, the readers often left these misreadings uncorrected, because they were paying more attention to the internal message they were generating than to the surface features of the text. For a person who knows both English and French it is usually irrelevant whether *door* or *porte* is what he hears or says; either word preserves the coherence of the message he is generating.

It is just this process of generating coherent messages from patterns of marks on a page that the skilled reader is engaged in. He is not, as one might think, involved in a piecemeal perception of individual letters and words. The process whereby the clues are selected and the messages are fashioned is one of the more challenging questions in the investigation of the way people process information. The challenge was put well by E. B. Huey, an outstanding early investigator of reading, when he said that "to completely analyze what we do when we read would almost be the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history."



DEFINITION OF READING is often extended to include matter presented in symbols other than an alphabet, including (*top to bottom*) music, logographic characters such as Chinese, a diagram of communication equipment, numerals and a map. Some of the techniques a reader brings to bear in reading a standard text presented in a familiar alphabet are also used for such symbols.

Escape Responses in Marine Invertebrates

Limpets, snails, clams, scallops, sea urchins and other slow-moving sea creatures go into remarkable gyrations when they are approached by a starfish. This lively behavior enables them to deter predation

by Howard M. Feder

One does not usually think of snails or other mollusks as being lively animals. Under certain conditions, however, some of these invertebrates exhibit a truly spectacular behavior. This was first noted many years ago by several marine biologists in Europe. Late in the 19th century Paul Schiemenz of Germany reported the curious response of a small sand-bottom snail (*Natica millipunctata*) to contact with a predaceous sea star, or starfish (*Marthasterias glacialis*). The snail would quickly extend a fold of its mantle tissue and slide it over its shell, thus dislodging any tube feet the starfish had planted on the shell. Other zoologists soon reported a variety of maneuvers used by mollusks to escape starfishes; some scallops were observed to begin violent swimming activities on being touched, and certain species of snails even executed a series of somersaults to shake off the predator.

These observations elicited little further research interest at the time. Then in the late 1940's Eugene C. Haderlie of the University of California at Berkeley reported on the "mild hysteria" and running behavior of the small mollusks known as limpets in response to contact with starfishes. This observation stimulated marine biologists to test a wide variety of mollusks for their reactions to sea stars. In a summary of these studies Theodore H. Bullock of the University of California at Los Angeles concluded that the elaborate forms of escape behavior exhibited by marine invertebrates must represent a phenomenon of considerable ecological importance. He suggested that these escape activities were not simply generalized reactions to contact but were specialized and apparently effective responses each species had evolved to cope with its predators.

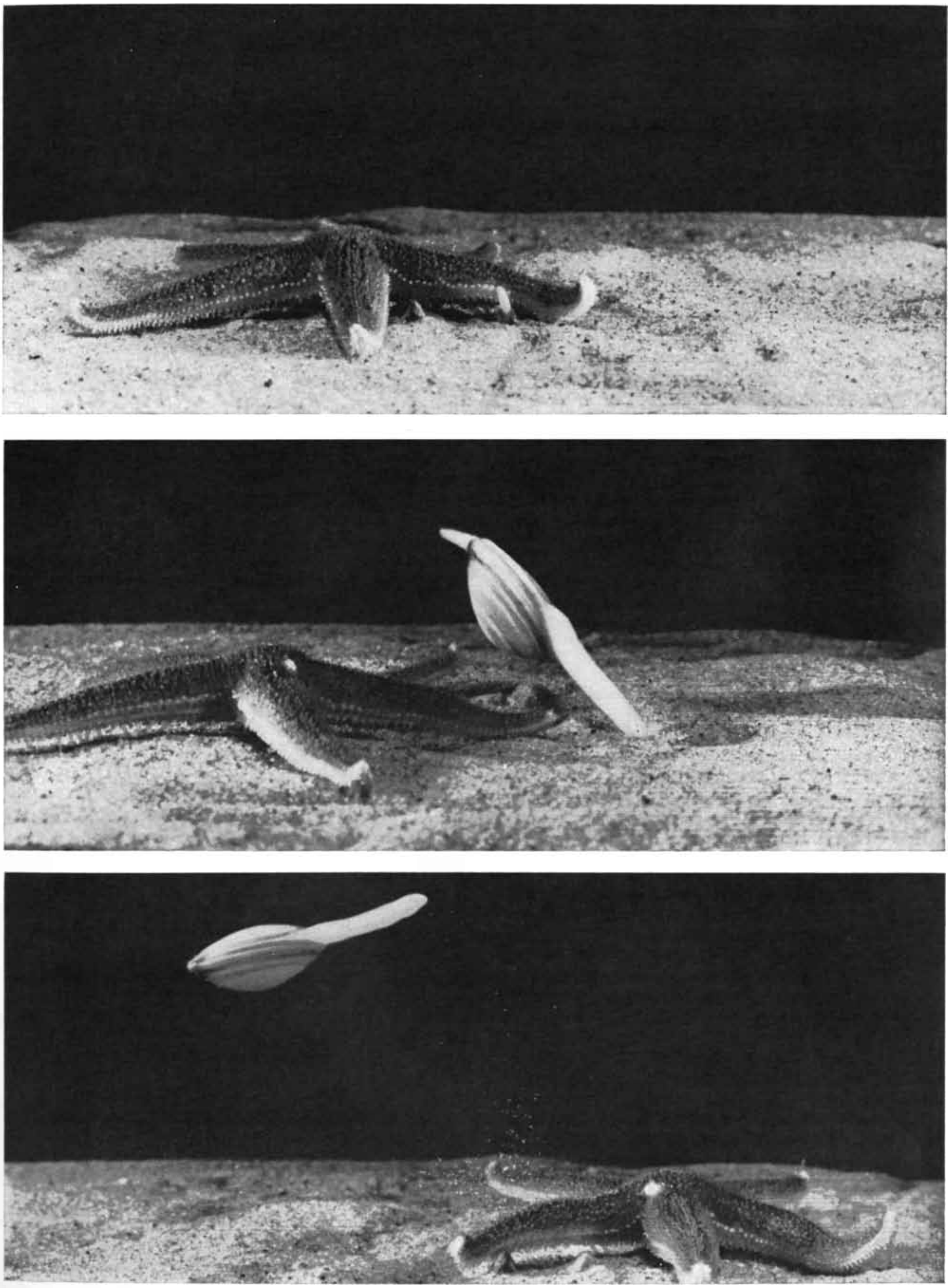
Currently there is a revival of interest in the investigation of this intriguing problem. The accumulation of reports from investigators around the world indicates that escape responses of prey species to predatory sea stars can probably be found wherever the investigator looks for them. How did the specific responses evolve? How effective are they in preserving the various species? What are the stimuli and mechanisms that produce the behavioral responses? These questions are being studied in a number of marine laboratories.

Sea stars inhabit most regions of the oceans and prey on many species of invertebrates. Most of the sessile victims (such as barnacles, mussels and other animals that are permanently attached to rocks or other substrates) manage to survive mainly by virtue of their high rate of reproduction. We are concerned in this article with the prey that have mobility: limpets, snails, clams, scallops, sea urchins, sea cucumbers, certain sea anemones and starfishes themselves (some species prey on other starfishes). All these animals have developed evasive maneuvers, some of them quite bizarre. Not all the responses depend on locomotion; in many instances the defense is a manipulative mechanism that enables the animal to remain in place while warding off the predator.

A survey of the escape behavior of such animals discloses several interesting points that invite detailed investigation. One is the variety of responses to the approach of the starfish. Another is the finding that related species show a great similarity in their escape behavior, even on opposite sides of the world. For instance, the running response that Haderlie observed in limpets in California is exhibited by limpets of the same genus in Europe as well as in North America, and the somersaulting movements characteristic of snails of the genus *Nassarius* are remarkably similar in related species in North America and in Europe. There are even more curious observations that need explanation. In some animals the supposed escape reaction initially includes a maneuver that would seem to make the animal more vulnerable to attack; for example, snails and limpets loosen their attachment to the underlying substrate immediately after contact. Furthermore, what are we to make of the finding that some of the animals that have highly efficient escape responses are apparently not attacked by starfishes at all?

Let us consider the varieties of escape behavior in some detail to see how effective they actually are for avoiding predation. Abe Margolin and I, working independently, closely examined the behavior of two different genera of limpets (*Diodora* and *Acmaea*) in the field and in the laboratory. (Margolin was at the Friday Harbor Laboratories of the University of Washington and I was then at the Hopkins Marine Station of Stanford University.) The two genera differ in their escape responses. When a limpet of the genus *Diodora* is touched by a species of starfish to which it is sensitive, the animal remains in place, raises its shell (in a movement that Bullock aptly called "mushrooming") and extends a

GREAT LEAP UPWARD is made by the cockle *Cardium echinatum* to escape from a preying sea star (*Asterias rubens*). The leap shown in the photograph on the opposite page is the last in a series of maneuvers executed by the cockle when it is touched by a sea star, beginning with extrusion of the tubular foot from the shell. The cockle continues jumping until it has managed to dislodge the starfish. The photograph was made by Holger Knudsen of the Marine Biological Laboratory at Helsingör in Denmark.



ESCAPE RESPONSE of a clam of the genus *Spisula* begins (*top*) when an arm of a sea star (*A. rubens*) touches the clam's siphon, which is protruding from the sand in this Knudsen photograph. The siphon is the small white object about a third of the distance in from the end of the right arm. On contact the clam stretches out its foot (*middle*) and thereby executes an escape jump (*bottom*).

fold of its mantle over the shell, thereby sweeping off the starfish's tube feet [*see illustrations on next page*]. The mushrooming reaction is so effective that *Diodora* is seldom attacked successfully by the two species of starfish whose intertidal distributions overlap the distributions of the limpets.

In contrast to *Diodora*, various species of *Acmaea* escape by running away, as Haderlie noted in his original observations of limpet behavior. Some species of *Acmaea* sense the approach of a predaceous starfish even before they are touched. After contact a limpet raises its shell in the mushrooming motion and rapidly moves away; its flight soon slows down, however, and the animal comes to a complete halt within a few minutes. In a laboratory aquarium, where room for flight is severely restricted, the limpets (and other organisms that depend on flight) are ultimately taken as food by predatory starfish in the same tank. In their coastal habitat, however, most species of *Acmaea* that can detect their predators are generally successful in evading capture. We have observed that the most important California intertidal starfish predator (*Pisaster ochraceus*) commonly feeds on only one species of *Acmaea,* known as the ribbed limpet; this limpet does not show an escape response. As an apparent consequence it thrives only in the high-tide region and its abundance declines sharply in the lower range where the starfish is active.

Of all the escape activities shown by animals attacked by starfishes perhaps the most spectacular are those exhibited by the snails of the genus *Nassarius*. These animals live on sediment bottoms and possess a strong foot that can flip the animal sideways or propel it into a somersault [*see illustration on page* 97]. Contact with a predatory starfish touches off a series of tumbling maneuvers by the snail, and the action is so violent that only a starfish considerably larger than the snail can remain attached to its prey. A large snail in New Zealand, *Struthiolaria papulosa,* displays a similarly striking performance. Robin Crump of the Orielton Field Centre in Wales, describing the behavior of *Struthiolaria,* reports that he has seen the animal perform as many as 50 consecutive somersaults in a four-minute encounter with a starfish. The European dog whelk (*Buccinum undatum*), a thick-shelled snail, also has a violent escape reaction. It rapidly rolls its shell counterclockwise and then with a twist of its foot throws the shell in the opposite direction, thereby effectively shaking off the starfish's tube feet.

The common black turban snail (*Tegula funebralis*) of intertidal waters in California has a varied repertory of responses after it has come in contact with a starfish. Contact causes it to flee, and if it is already in motion, it will speed up its travel from its normal two or three centimeters per minute to eight centimeters per minute. When an anterior surface of the snail is touched, it raises the front portion of its foot, makes a turn of about 90 degrees and moves off in the new direction. A touch on its side causes the animal to twist its shell away from the point of contact; when a posterior surface is touched, the snail tilts the shell over its head, twists about violently and moves off rapidly. On a steep slope the starfish contact will precipitate a tumble by the snail down the slope. In many areas black turban snails are apparently not fed on by sea stars in proportion to their abundance, and generally they are not attacked to any extent if sessile organisms such as barnacles or mussels are available.

The abalones (genus *Haliotis*) are another group of shellfish whose members exhibit very effective escape maneuvers. When touched by a starfish, an abalone raises its shell and whirls it violently from side to side; some species are capable of rapid locomotion, lifting the forward part of the foot and extending it far ahead for a leaping maneuver. The black abalones (*H. cracherodii*), dwelling in intertidal areas, seldom fall prey to starfishes. The predominantly subtidal red abalones (*H. rufescens*) are also usually successful in escaping the predators (except for the small young individuals, which can readily be taken by the large sunflower sea star *Pycnopodia helianthoides*).

Long before biologists began to notice the escape behavior of mollusks the Maori people of New Zealand were well acquainted with this behavior in abalones and put it to everyday use: they harvested abalones (a prized food) from inaccessible locations by deliberately touching them with a predatory sea star and then plucking the rapidly moving mollusks from the substrate. David Montgomery of California State Polytechnic College has discovered an instance in which an abalone, under attack by a starfish, liberates a substance that serves to alarm other members of its own species. When it is touched by a starfish, the animal exudes a cloudy fluid that causes other abalones in the vicinity to make escape movements typical of the species' reaction to contact with a sea-star predator. It was already known that some freshwater gastropods and tropical sea urchins discharge substances signaling fright or injury to members of their own species. The same phenomenon is well known in the world of the insects, some of which, notably certain ants, emit alarm pheromones [see "Pheromones," by Edward O. Wilson; SCIENTIFIC AMERICAN, May, 1963].

Even among the bivalve mollusks (clams, cockles, scallops), most of which would seem to have scant capability for rapid locomotion, one finds remarkable feats of agility. Typically a bivalve attacked by a sea star remains in place and slowly closes its valves (shells). Some bivalves show little sensitivity to the starfishes that prey on them; in fact, a few open their shells and continue to feed while an attacking starfish is manipulating them. Other bivalves, however, have evolved not only vigorous escape responses but also the ability to recognize their predators. Scallops can swim or jump about by rapidly opening and closing their shells; such movements, although uncontrolled as to direction, apparently are rather effective in protecting these mollusks from extensive predation. The cockles and a few species of clams use a different mechanism: when touched by a starfish, such a bivalve opens its shells wide, thrusts out its foot against the substrate on which it is resting and with a violent push of the foot makes a leap upward, often as far as 10 centimeters [*see illustrations on page 92 and on opposite page*].

The sea urchins, members of the phylum Echinodermata with a globular form studded with spines, possess two methods of protection against starfishes. They can run or use poisonous pinching structures, called pedicellariae, to grasp and remove the starfish's probing tube feet. In the latter case contact by the sea star generally results in a rapid retraction of the sea urchin's tube feet, a lowering of its spines and a gaping and erection of the pedicellariae. Significantly, a European species of urchin that customarily occupies exposed positions on rocks has sufficient speed to outdistance a pursuing starfish, whereas two slower European species of urchins typically inhabit protected niches among the rocks.

Another species, the purple urchin of North American coastal waters (*Strongylocentrotus purpuratus*), often aggregates in large intertidal groups, making up formidable beds of thousands of individuals. Although the urchins lie exposed to starfishes, the carpet of pinching pedicellariae they present undoubt-

EXTRUDED MANTLE is the response of the keyhole limpet *Diodora aspera* to contact with the sea star *Pisaster ochraceus*. Extrusion of the soft mantle can be seen most clearly in the limpet at right. The reaction, which is shown in more detail in the illustration at the bottom of this page, gives the limpet's shell a fleshy covering. The tube feet on the arm of the predatory sea star cannot grasp the covering firmly and may in fact be repelled by it.

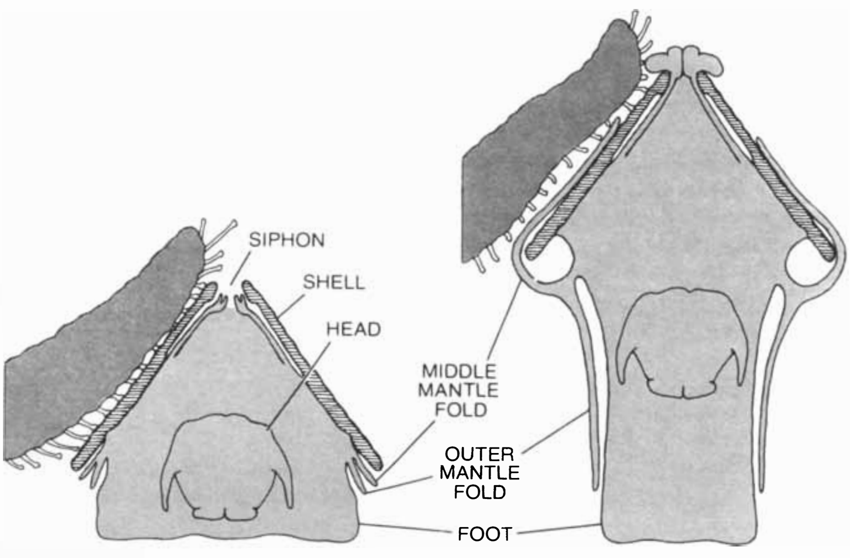
edly deters attack [*see illustrations on page* 99]. Individuals of this urchin species also escape from predatory sea stars by means of a running response if the encounter between the two takes place on a uniform substrate. Recent work done on subtidal populations of the purple urchin off Point Loma, Calif., by Richard J. Rosenthal of the Scripps Institution of Oceanography and James R. Chess of the Tiburon Fisheries Laboratory of the National Marine Fisheries Service has shown that here the leather star (*Dermasterias imbricata*) is a major predator on urchins. This sea star has not been reported to prey on urchins elsewhere along the northern Pacific coast. Rosenthal and Chess suggest that the unusually high level of predation observed off Point Loma represents a regional phenomenon. There seems to have been a population explosion among the sea urchins in an area practically devoid of the organisms typically fed on by the leather star elsewhere, and the sea star has apparently switched to an abundant but less "preferred" prey species.

Certain other echinoderms show escape reactions in response to a stimulus from predatory starfishes. A large sausage-shaped animal (*Parastichopus californicus*) known as a sea cucumber responds with awkward swimming motions produced by contractions of its longitudinal muscles. And among other invertebrates exhibiting sensitivity to the starfishes are some of the sea anemones, flower-like members of the phylum Coelenterata with a bush of tentacles topping a thick column. By bending the column anemones (*Stomphia* and *Actinostola*) make creeping or swimming movements to escape from a predator.

Several species of sun stars (*Solaster* and *Crossaster papposus*) feed intensively on stars of other genera and sometimes even attack smaller individuals of their own species; such small prey are readily captured and are ingested completely. All the prey species move away from the attacking sea stars after contact with them, and large stars may outdistance the pursuing predators. Karl P. Mauzey and his associates, who formerly worked at the University of Washington, noted that predatory sun stars can often counteract the rapid escape movements of their potential prey. The predator typically moves along alternately raising and lowering its leading arms. When the tube feet on one of the arms come in contact with another sea star, the arm drops onto the upper surface of the prey. The predator then raises the other arms facing the prey and "lurches" forward onto it. If the prey is the sun star *Solaster stimpsoni*, however, it may still avoid capture by curling its arms back over the upper surface of its body and using these arms to push upward and backward against the attacker. When a predatory sun star attacks a large sea star of the European genus *Asterias*, the captive often frees itself by throwing off its arms; the predator then feeds on these amputated members. Occasionally a sun star smaller than its prey has been seen clinging to the victim and feeding on its arm tips. In a laboratory tank the predator sometimes will cling tenaciously to a captive and feed in this way for several days.

So much for the observations of escape activities. Attempts to explain their development and their role in stabilizing animal communities in the sea begin with the examination of apparent anomalies. One of these has been the puzzling circumstance that many of the animals that show escape responses to starfishes in the laboratory are rarely, if ever, fed on by these predators in nature and in some cases apparently do not even inhabit the same waters.

Mauzey and his associates, whose investigations included free-diving observations of marine populations, suggest

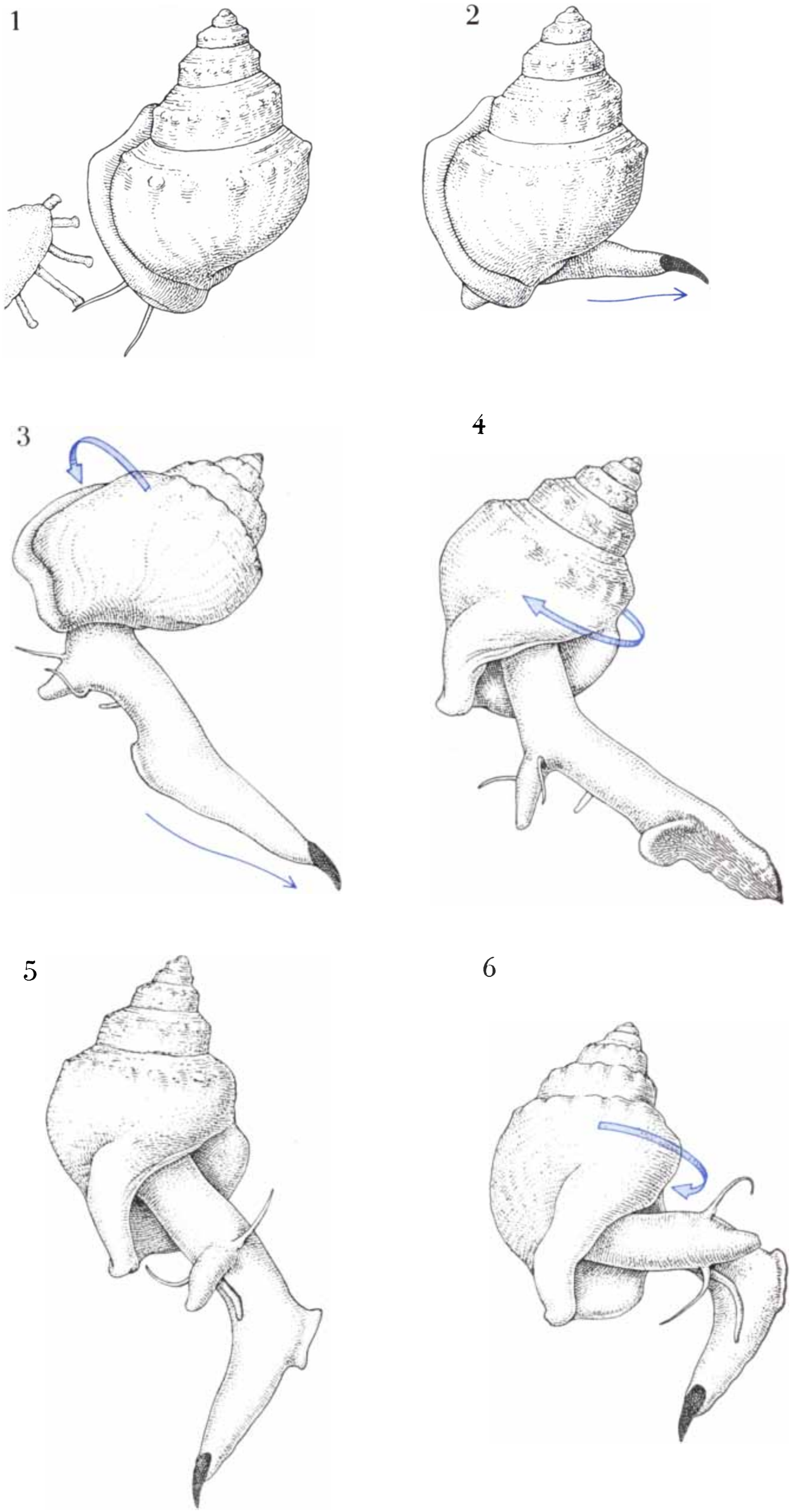


POSITIONS OF LIMPET are portrayed in the normal state (*left*) and when the animal is reacting to a sea star (*right*). The limpet is usually attached firmly to a rock by its foot. In the mantle response the animal stretches upward and then slides its mantle over its shell.

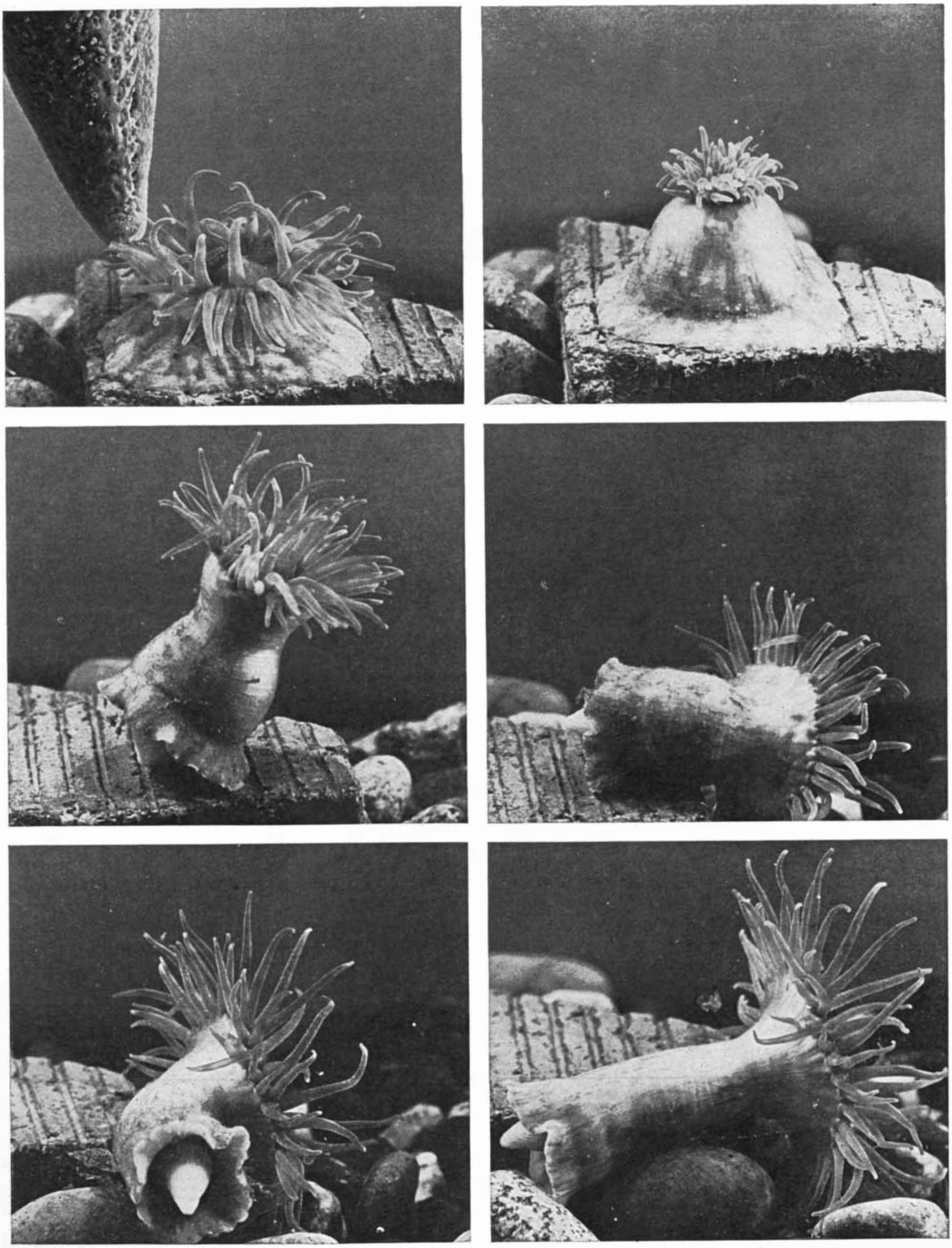
that the unmolested animals may actually owe their freedom from attack to the almost total effectiveness of escape reactions; such reactions may completely remove the potential prey from the predators' diet. Mauzey's group cites certain observations that tend to support this view.

One of the immune organisms is the large sea cucumber *Parastichopus californicus*. Because starfishes had never been seen to attack this animal in nature, observers supposed that the cucumber's reactions to starfishes in the laboratory did not actually represent escape responses related to this predator. Recently, however, intensive searching has turned up a few cases of successful predation on *Parastichopus* by the sunflower starfish, and it has also been found that all the other starfish species to which the sea cucumber responds with swimming movements feed on other species of sea cucumbers. Hence it appears likely that *Parastichopus*' activity was indeed developed as a defense against starfish predators and that the almost complete lack of predation on the large sea cucumber results from the effectiveness of the response. The other observations cited by Mauzey and his colleagues to support their hypothesis has to do with certain species of sea anemones. Under laboratory conditions these anemones react to contact with four species of starfish. There was reason to doubt that this reaction actually represented a predation-induced escape response, because in nature these anemones and starfishes had not been found to occur together in the same waters. In diving explorations Mauzey's group did, however, eventually find some instances of co-occurrence, and the starfishes there were feeding on the anemones. The four starfish species mentioned have also been observed to feed on other species of anemones and are almost the only sea stars with such a diet. It seems reasonable to conclude that the evasive behavior shown by the various anemone species is truly a response to the predatory sea stars.

What is the stimulus that initiates an escape reaction on the part of the prey of starfishes? It is not necessarily contact with the predator. Generally the most vigorous escape movements occur only after contact, but early in the investigation of the phenomenon Bullock and I discovered that many gastropods reacted to the approach of a starfish even before contact. This suggested that the animals responded to substances emanating from the predator. Investigation



SEQUENCE OF MOVEMENTS in the rolling flight response of the New Zealand snail *Struthiolaria papulosa* following contact with a sea star (*Astrostole scabra*) is depicted. When a tube foot of the star touches a tentacle of the snail (*1*), the snail's foot extends to the side of the shell (*2*) and the shell is rolled by violent twisting of the body (*3*). Eventually the snail becomes inverted (*4*). Then the foot begins to curl under shell (*5*), soon attaining a position from which it can throw the shell in the opposite direction (*6*). Illustration is derived from a paper by Robin Crump of the Orielton Field Centre in Wales.



AVOIDANCE MOVEMENTS are made by the sea anemone *Stomphia coccinea* on contact with the sea star *Dermasterias imbricata*. At top left the star touches the anemone and the animal withdraws its tentacles (*top right*), meanwhile beginning to detach itself from the substrate. A few seconds later (*middle left*) the pedal disk has been detached from the substrate and a conelike structure is beginning to form at the bottom of the animal. Soon (*middle right*) the animal moves by flexing its body, and within about 10 seconds (*bottom left*) it has moved off the substrate. Finally (*bottom right*) the animal relaxes; it will resettle in about 15 minutes.

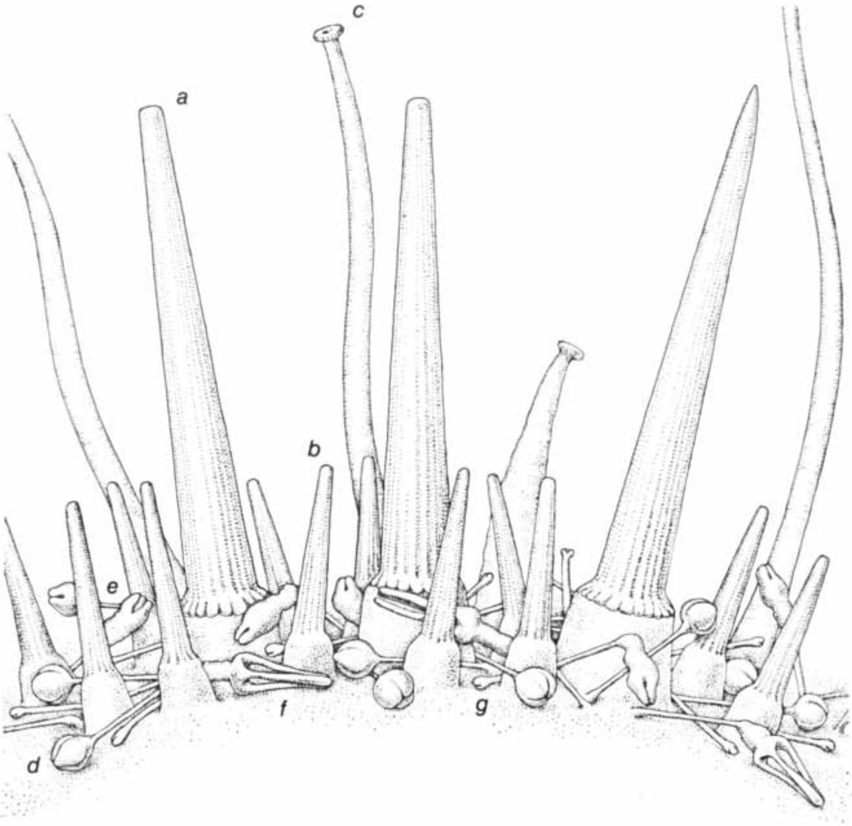
has since established that an active, irritating material is in fact continuously liberated by predatory starfishes. These substances are found in most tissues of the starfish; primarily, however, they are concentrated in the epithelial covering of the body, particularly in the tube feet.

Drippings exuded by starfishes and crude extracts from starfish tissues have been shown to be capable of stimulating escape responses in their prey. All the predatory starfishes of North America and Europe that have been examined possess active substances. The extracts are all relatively thermostable, weakly dialyzable, insoluble in fat solvents, attached to or readily absorbed by proteins and stable for many months when frozen or freeze-dried.

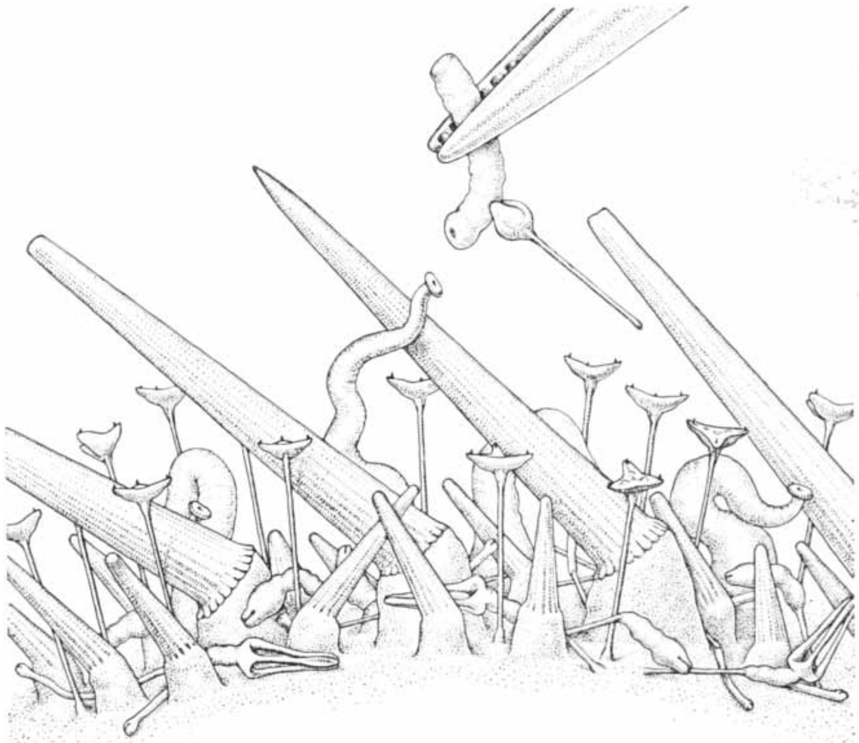
Sea-star drippings and crude extracts contain steroid saponins that are known to be toxic, and it has already been demonstrated in at least one case that saponin-like substances are responsible for a mollusk's avoidance reaction to a starfish; this was shown by Alexander M. Mackie, Reuben Lasker and Patrick T. Grant in the response of the dog whelk to the sea star *Marthasterias glacialis.* (Mackie, Lasker and Grant were working at the University of Aberdeen in the Fisheries Biochemical Research Unit of the British Natural Environment Research Council.) The similarity of the properties of the active materials obtained from the predatory starfishes so far examined suggests that perhaps all the predators contain saponin-like substances.

The indication that all predatory starfishes emit chemically similar irritating substances may explain the worldwide reactions to these animals by marine invertebrates. It could also account in part for the observation that often prey organisms respond with avoidance reactions even to species of starfish that they never normally encounter or that do not normally feed on them in nature. Mauzey and his colleagues point out that some of the anomalous responses might also be explained if one assumes that some of the chemical irritants originate in prey organisms and that the sea star which atypically induces avoidance responses has a diet similar to the diet of the predator.

In studies at laboratories in the U.S. and in Sweden, Montgomery, Jan Arvidsson and I have obtained evidence suggesting that most and possibly all echinoderms contain and exude substances that can induce avoidance reactions. Even nonpredatory echinoderms (sea urchins, sea cucumbers, and star-



UNDISTURBED SURFACE of the sea urchin *Psammechinus miliaris* is depicted. The anatomical features include primary spines (*a*), secondary spines (*b*), tube feet (*c*) and several kinds of pedicellariae (*d–g*). The primary spines are about five millimeters long.



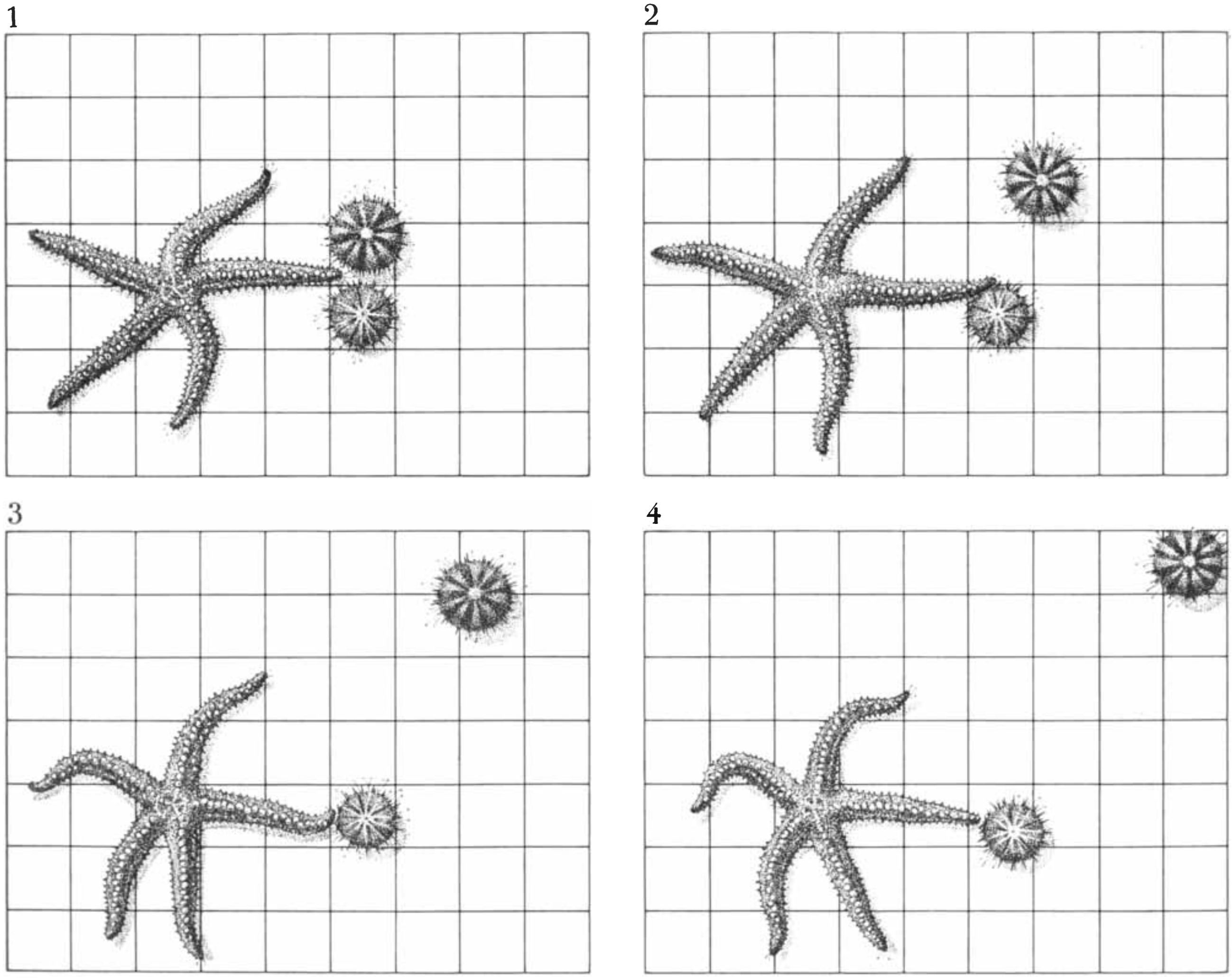
DEFENSIVE REACTION is manifested on the surface of *P. miliaris* when a tube foot of the sea star *Marthasterias glacialis* is held nearby with forceps. The spines bend away, and globiferous pedicellariae are raised and opened to close on the foot. Illustrations on this page are based on drawings by Kai Olsen of Marine Biological Laboratory at Helsingör.

fishes that are not predatory) contain irritating materials; however, they do not liberate any significant amount of it, either because their tissues contain very little of it or because it is held in a tightly bound form. In addition many nonpredatory echinoderms contain saponins. Therefore it is possible that steroid-like irritating substances are common metabolites in all echinoderms but that these substances are released in detectable amounts only by predatory sea stars.

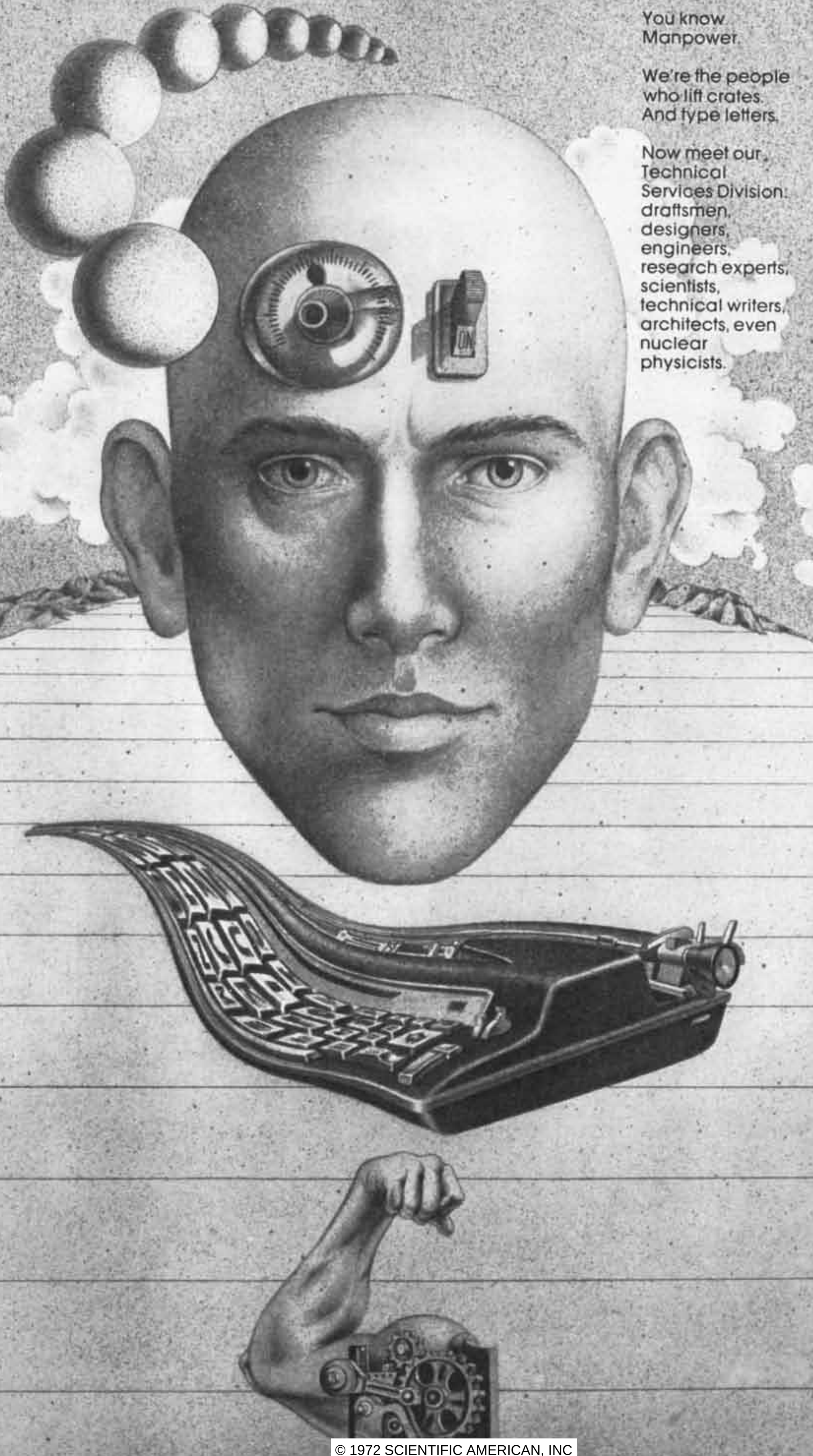
How did it come about that predatory starfishes evolved the practice of exuding substances that alarm their prey? There is reason to believe that this behavior may have served originally as an effective prey-catching device. As we have noted, some species of limpets and abalones respond to a starfish's approach by raising their shells and loosening their attachment to the underlying substrate. Such a response, induced by an irritating toxic substance (which might also serve the sea stars as a repellent against predators), may initially have enabled the predator to carry off the prey more easily. The prey organisms in turn may have evolved devices for fleeing from or warding off the predator, presumably as responses to the chemical stimulus. Jefferson Gonor of Oregon State University has suggested that most of the escape maneuvers of marine snails, for example, are probably not new acquisitions but instead evolved from existing locomotory and righting movements that originally served functions other than defense.

The most novel and intriguing development, in Gonor's view, is the prey animals' acquisition of the ability to identify and recognize their specific predators. In most cases examined by our group and by other investigators a given starfish predator elicits strong responses only from the animals that serve as its prey. For example, D. Craig Edwards of the University of Massachusetts has shown that the olive snail responds strongly to a starfish species that preys on it but does not react significantly to another species of the same genus that it never encounters in nature. Similarly, the purple sea urchin shows almost no response to starfishes that rarely feed on it, and anemones do not swim away when they come in contact with starfishes that do not prey on coelenterates.

Further investigation of the chemical communications among all these animals should lead us to a clearer view of life on the sea bottom and at its teeming margins.



SEA URCHINS display differing reactions to a sea star depending on whether or not the urchin's natural habitat affords protection from sea stars. In a laboratory experiment the animals were placed on a grid with lines five centimeters apart. The urchin *Psammechinus miliaris*, which is from the protected habitat, barely moves during a period of 90 seconds, whereas the urchin *Strongylocentrotus droebachiensis*, which is from the unprotected habitat, moves away rapidly. The sea star in the experiment is *M. glacialis*.



MATHEMATICAL GAMES

Amazing mathematical card tricks that do not require prestidigitation

by Martin Gardner

"Do you like card tricks?"
"No, I hate card tricks," I answered.
"Well, I'll just show you this one."
He showed me three.
—SOMERSET MAUGHAM, *Mr. Know-All*

Maugham's experience with card magicians is all too familiar. "I don't really like people who do card tricks," Elsa Maxwell once wrote (I quote from an autobiography of a lady magician, *You Don't Have to Be Crazy,* by Frances Ireland). "They never stop at one or two, but go on and on and on, and always make you take cards, or turn up cards, or cover cards, until you are worn out."

Mathematical card tricks, let it be admitted at once, are precisely the kind of tricks that are the most boring to most people. Nevertheless, they have a curious appeal to mathematicians and mathematically minded magicians. Hundreds of new card tricks are published every year in magic periodicals, particularly in *The Pallbearers Review.* This is a four- to eight-page monthly available for $6 per year from systems engineer Karl Fulves, P.O. Box 433, Teaneck, N.J. 07666. It has been said, although Fulves denies it, that the title is in ironic reference to the widespread opinion among professional prestidigitators that "magic is dead." (I must warn any prospective subscriber that he will find the explanations of most tricks unintelligible if he is not familiar with the rudiments of magic.)

Many excellent card deceptions are based on a parity principle, but the underlying even-odd structure is usually concealed so ingeniously that if you follow the directions with cards in hand you are likely to astonish yourself. Consider the following simplified version of a trick invented by George Sands and published some 20 years ago in a magic magazine called *The Genii.* Magicians classify it as an "oil and water" effect, for reasons that will be apparent in a moment. There are many ways of achieving the same effect by secret and difficult "moves," but this version is entirely self-working.

Remove 10 red and 10 black cards from the deck and arrange them in two face-up piles, side by side, with all red cards on the left and all black cards on the right. First you tell your watchers that you will demonstrate what you intend to do by using only five cards of each color. With both hands simultaneously remove the top card from each pile and place them, still face up, on the table at the bottom of each pile. Do the same with the next two top cards, but this time cross your arms before you place the two cards on the two new piles you are starting. This puts a black card on the red one and a red card on the black one. The next transfer of a pair of cards is made with uncrossed arms, the next with crossed arms, and the fifth and last pair is dealt with arms uncrossed. In other words, five simultaneous deals are made, with arms crossed only on alternate deals. On each side you now have a pile of five face-up cards with their colors alternating. Put either pile on the other one. Spread the 10 cards to show that colors alternate throughout.

Square the cards and turn the packet face down. From its top deal the cards singly and face up to form two piles again, dealing alternately to the left and right. Call attention to the fact that this procedure naturally separates the colors. At the finish you will have five reds on the left and five blacks on the right.

State that you will repeat this simple series of operations with all 20 cards. Begin as before, with 10 face-up reds on the left and 10 face-up blacks on the right. Transfer the cards to form two new piles, just as you did before, crossing your arms on alternate deals so that the colors alternate in each pile. After all 20 cards are dealt put one pile on the other, square the cards, turn the packet over and hold it face down in your left hand.

Deal 10 cards face up to form two piles, dealing from left to right and observing aloud that this brings the reds together on the left and the blacks together on the right. After the 10 cards have been dealt face up do not pause but continue smoothly and deal the remaining 10 cards face down. It is best to put down the cards so that they overlap in two vertical rows [*see illustration on opposite page*].

Pick up the five face-down cards on the left with your left hand and the five face-down cards on the right with your right hand. Cross your arms and put the cards down. You explain that you have transferred half of the cards of each pile to the pile of the opposite color but that like oil and water the colors mysteriously refuse to mix. Turn over the face-down cards. To everyone's surprise (you hope) the reds are back with the reds and the blacks are back with the blacks! Readers should have little difficulty discovering why it works with any set of cards containing an even number of cards of each color and why it did not work when you demonstrated it with 10 cards.

After you have finished the oil-and-water trick put the two piles together with either color on top. Turn the packet face down and spread it in a fan. You are ready to perform a red-black trick invented by Fulves and published in his review in September, 1971.

Ask someone to pull slightly forward any 10 cards he pleases. The fan will resemble the one shown in the illustration on page 104. With your right hand count the jogged (protruding) cards to make sure there are 10. Do this by removing the cards one at a time from right to left, putting them into a face-down pile as you count from one to 10. Close up the 10 cards remaining in your left hand and place them in a second face-down pile alongside the first.

Tell your audience that an amazing thing has happened. Although 10 cards were selected randomly, the colors in the two piles are so ordered that every *n*th card in one pile has a color opposite to the color of the *n*th card in the other pile. To prove this, turn over the top cards of each pile simultaneously. One will be red and the other black. Place the black under the red, turn the pair over and put it aside to form a new face-down pile. Repeat the procedure with the cards now on top of the two original piles. They will be red-black too. Indeed, every pair you turn will be red-black!

As you show the pairs always put the black card under the red before you turn them over and place them on the third pile. When you finish, the cards in this face-down pile will have alternating colors.

Now you are ready to perform a truly mystifying trick in which parity is conserved in spite of repeated shuffling. Known as Color Scheme, it was invented by Oscar Weigle, an amateur magician who is now an editor at Grosset & Dunlap. It sold as a manuscript in magic stores in 1949.

Give the packet of 20 cards to someone and ask him to hold it under the table where neither he nor anyone else can see the cards. Tell him to mix the cards by the following procedure. (It is known as the Hummer shuffle, after Bob Hummer, the magician who first used it in tricks.) Turn over the top two cards (not one at a time but both together as if they were one card), place them on top and cut the packet. Your assistant is to keep repeating this procedure of turn two, cut, turn two, cut for as long as he wishes. The procedure will, of course, result in a packet containing an unknown number of randomly distributed reversed cards.

With the cards still held under the table, tell your assistant to do the following. Shift the top card to the bottom. Then turn over the next card, produce it from under the table and place it on the table. This procedure is repeated—card to bottom, reverse next card and deal—until 10 cards have been dealt to the table. It will be apparent that the cards have become mysteriously ordered. All the face-up cards are the same color and all the face-down ones are of the opposite color.

The second and climactic half of the trick, which Weigle confesses is a "barefaced swindle," now unfolds. Your assistant is still holding 10 cards under the table. Ask him to shuffle them by separating them into two packets; then, keeping all the cards flat (no card must be allowed to turn over), weave the two packets into each other in a completely random way. You can demonstrate how to do this by using the 10 cards already dealt. After your assistant has executed the shuffle a few times, ask him to turn over the packet and shuffle the same way a few more times. If he likes, he can give the packet a final cut.

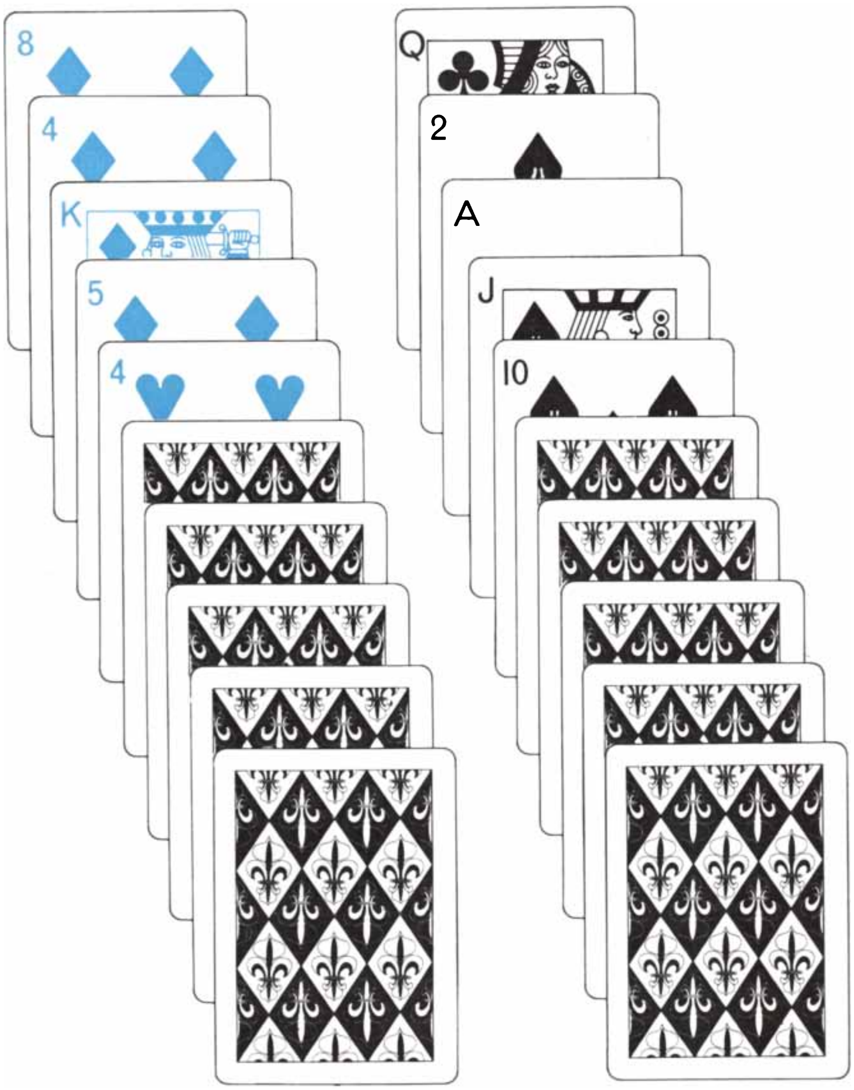
Now he continues with the dealing procedure he used before: card to bottom, next card reversed and dealt. (The final card is reversed and dealt.) In spite of the thorough mixing the result is exactly the same as before. All the face-up cards match the former face-up cards in color, and the same is true of all face-down cards.

One of the oldest themes in card magic is to produce in some startling fashion a card that has been randomly selected and replaced. Here is a simple method that exploits a binary sorting technique. Fulves published it in his review in November, 1970.

Take 16 cards from a shuffled deck and spread them face down on the table without mentioning how many cards you are using. A viewer selects a card, looks at it and places it on top of the deck. The remaining cards in the spread are squared and put on top of the deck above the chosen card. Ask him to cut off about half of the deck, give or take half a dozen cards. Actually he can take between 16 and 32 cards. He hands this packet to you.

Hold the packet in both hands. As your left thumb slides the cards one at a time to the right, move your right hand forward and back so that every other card, starting with the first one, is jogged forward. The resulting fan of cards will resemble the one in the illustration on the next page except that the jogged cards are not randomly distributed. Strip all the projecting cards from the fan and discard them. Square the remaining cards and repeat the procedure, jogging forward all the cards at odd positions, starting with the first card. Strip them out and discard. Continue in this way until only one card is left. Before turning it face up ask for the name of the selected card. It will be the card you are holding.

A completely different method of locating a selected card is explained in a



The oil-and-water effect

1968 magic book by Eddie Field and Jon Racherbaumer, *The Artful Dodges of Eddie Field.* Turn your back and instruct someone to cut a shuffled deck into three approximately equal piles. He turns over any pile and then reassembles the deck by sandwiching the face-up pile between the other two, which remain face down. He is told to remember the top card of the face-up pile. With your back still turned, ask him to cut the deck several times, then give it one thorough riffle-shuffle. The shuffle will of course distribute the face-up cards randomly throughout the deck.

Turn around, reverse the pack and spread it in a row. Look for a long run of face-up cards, remembering that a cut may have split the run so that part of it is at each end of the spread. The first face-down card above the run is the chosen one. Slide it from the spread, have the card named and then turn it over.

Our last trick, based on a curious shuffling principle discovered by Fulves, is presented as a gambling proposition. All cards of one suit (the suit can be chosen by the victim) are removed from the deck. Assume that the discarded suit is diamonds. The remaining cards are arranged so that each triplet has three different suits in the same order. (Card values are ignored.) Again the victim may specify the ordering. Suppose he chooses spades, hearts and clubs. The 39-card deck is arranged from the top down so that the suits follow the sequence spades, hearts, clubs, spades, hearts, clubs and so on.

Place the deck face up in front of the victim. Ask him to cut it in two packets and riffle-shuffle them together. As he makes the cut, note the suit exposed on top of the *lower* half. We shall call this suit k. After the single shuffle the deck is turned face down. The cards are now taken from the top three cards at a time, and each triplet is checked to see if it contains two cards of the same suit.

It is hard to believe, but:

(1) If k is spades, no triplet will contain two spades.

(2) If k is hearts, no triplet will contain two clubs.

(3) If k is clubs, no triplet will contain two hearts.

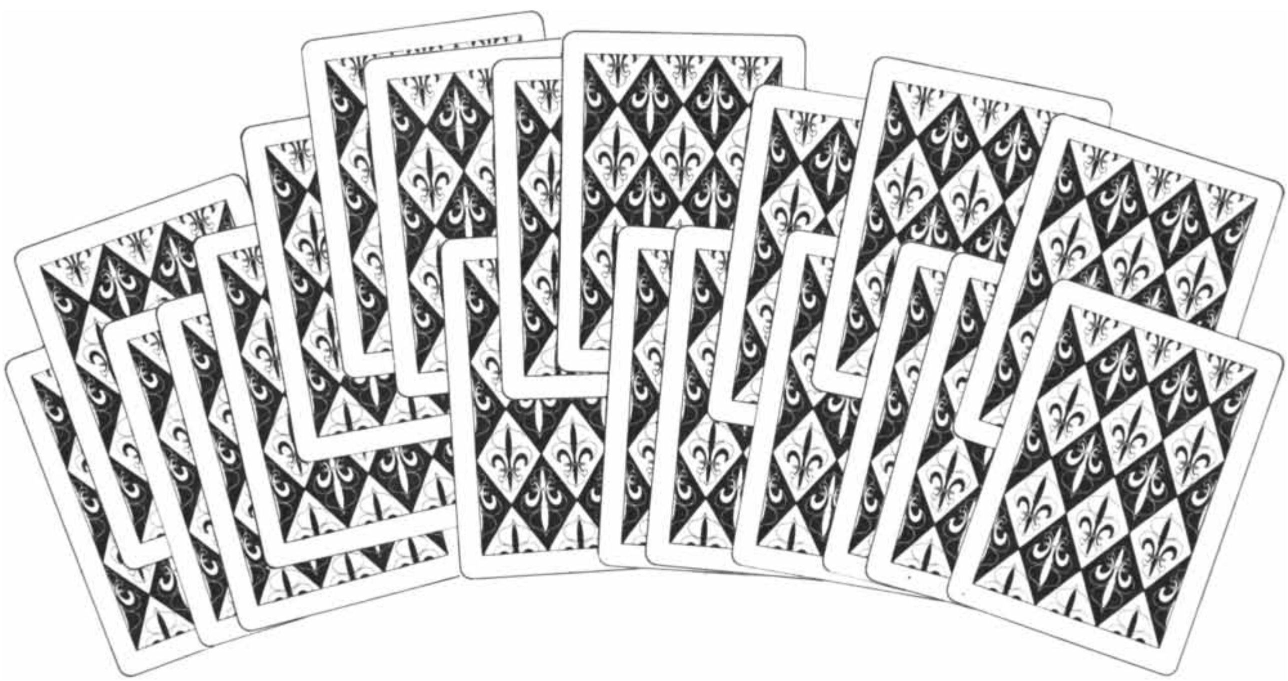
This assumes, of course, a spades-hearts-clubs ordering. If the ordering is otherwise, the three rules must be modified accordingly; that is, spades must be changed to whatever suit is at the top of each triplet, and so on. Let m stand for the suit that you know cannot show twice in any triplet, and a and b for the suits that can.

Before dealing through the deck to inspect the triplets, make the following betting proposition. For every triplet containing a pair of m's you will pay the victim $10. In return he must agree to pay you 10 cents for every pair of a's or b's. It seems like a good bet for the victim, but it is impossible for you to lose, and the swindle can be repeated as often as you please. Just arrange the cards again and allow the victim to make the single riffle-shuffle. Naturally you always promise to pay him for doublets of the suit that you know cannot show. The fact that this suit may vary from deal to deal makes the bet particularly mystifying.

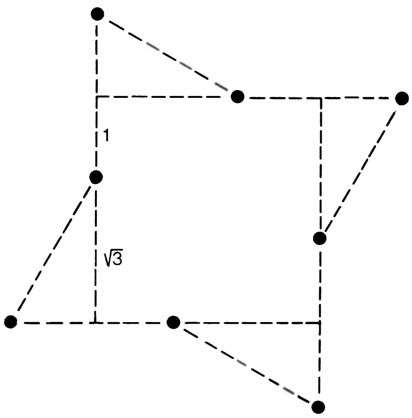
As Fulves has observed, the triplets have other unexpected properties. Of the triplets containing pairs the a's and b's will alternate; after a pair of a's the next pair will be b's and vice versa. Pairs of one suit always include a top card of the triplet. Pairs of the other suit always include a bottom card.

No explanation of these tricks will be given next month. Readers will find it stimulating, however, to analyze each trick to see if they can comprehend exactly why it operates with such uncanny precision.

So many readers proved that May's solution to the nine-digit problem ($532 \times 14 = 98 \times 76 = 7,448$) did indeed give the highest product that it is impossible to list their names. The problem has 11 basic solutions, starting with $158 \times 23 = 79 \times 46 = 3,634$ as the one with the lowest product. All 11 were found by hand by Ralph G. Beaman, Emmet J. Duffy, Matthew Hodgart and Allan L. Sluizer, and by computer programs written by L. M. Arnett, Philip Coates, Mark Dulcey, R. F. Forker,



Ten cards jogged forward



The solution to last month's problem

Charles M. Hains, Richard Hendrickson, Jeffrey Hoel, Mohan S. Kalelkar, Stephan Ketcham, Daniel S. Marcus, Gerald A. Mischke and William V. Snyder.

Victor Meally wrote and asked what the maximum product is when all 10 digits are plugged into the expression ABC $\times$ DE = FGH $\times$ IJ. His solution ($915 \times 64 = 732 \times 80 = 58{,}560$) was confirmed independently by Hendrickson, Sluizer and Forker, who found 64 solutions, assuming that 0 is excluded as an initial digit. The solution with the smallest product is $306 \times 27 = 459 \times 18 = 8{,}262$.

The seventh chessboard illustration in May had three men incorrectly colored. On White's seventh row the three men on the right should be a white king and knight and a black pawn. The pawn on QB2 in the ninth picture should be on Q2, and in the fifth picture a black queen belongs on the empty border cell. Some readers counted only 312 capture moves in the eighth picture, forgetting that a promoted pawn can become any of four pieces, each choice a different "move."

Improvements have been reported on two of the chess tasks. Robert T. Wainwright found a way of placing 10 queens on the order-10 board so that 22 (not 21) vacant cells were unattacked. Thur Row writes that in 1952 the maximum number of moves for a legal position, with no promoted men or promotion moves, was raised from 178 to 181, using 22 men, and that he recently found a way to get the same total with 21 men.

The solution for last month's problem of placing eight points so that the perpendicular bisector of each pair of points passes through at least two other points is shown in the illustration above. Next month I shall comment on letters received concerning answers to the other short problems given in April.

THE AMATEUR SCIENTIST

An amateur makes a wind tunnel to study the vortexes that form around a cylinder

Conducted by C. L. Stong

The flow of air around a cylindrical object in a direction perpendicular to the axis of the cylinder tends to generate vibrations. The ancients applied the effect in a musical instrument known as the aeolian harp, which consists of a set of strings stretched across a sounding box. The strings emit a series of harmonic tones that vary with the strength of the breeze.

The aeolian effect was investigated in some depth by experimenters of the 19th century, although the aerodynamic mechanism that is responsible for the vibrations is not fully understood even today. Interest in the vibrations has risen sharply in recent years. They can generate destructive forces in cylindrical structures ranging from smokestacks and water towers to rockets on launching pads. Recently Bonnie Jean Luessen, a high school student in Huntsville, Ala., built a small wind tunnel for investigating the aeolian effect quantitatively. The project placed her among the finalists in the 31st annual Science Talent Search. She describes her experiments as follows:

"When a stream of air flows around a slender cylinder at relatively low velocity, two stable, symmetrical vortexes form in the wake. At a somewhat higher velocity one of the vortexes becomes unstable, detaches from the cylinder and moves downstream. The action transmits a force through the air that disturbs the remaining vortex, which thereupon becomes unstable in its turn, detaches and drifts downstream. A new vortex forms at the site of the first disturbance. Thereafter vortexes form and detach alternately as long as the critical velocity persists.

"The swirling pattern of flow in the wake of the cylinder is known as von Kármán's vortex street in honor of the Hungarian aerodynamicist Theodor von Kármán, who explained the effect mathematically some years ago. The periodic shedding of vortexes of opposite rotation generates alternating forces that act on the cylinder at right angles to the flow of air. The influence of the oscillating forces increases as the length of the cylinder increases and as the diameter decreases. The amplitude of vibration also increases as the frequency of the alternating forces approaches the natural period at which the cylinder vibrates.

"Wind speeds of less than 30 knots induced significant vibrations in the first Redstone rocket to be erected on its pad. The problem was solved by sheltering the vehicle from winds of more than 25 knots. Wind-tunnel tests subsequently indicated that the *Saturn V* vehicle would be susceptible to wind-induced oscillations at velocities occasionally observed at Cape Kennedy. To prevent the buildup of destructive forces the upper end of the rocket assembly was clamped to a bracing structure through a pair of hydraulic damping cylinders similar in principle to the shock absorbers of automobiles. The cylinders dissipate the wind-induced energy as heat.

"My experiment was designed for ob-

The form of vortexes around a cylinder in a wind tunnel serving the aeolian effect in a homemade wind tunnel. In particular I set out to investigate von Kármán's vortex street by measuring the dynamic responses of the cylinder instead of injecting smoke in the wake to make the vortexes visible, a procedure that is often followed. Experiments done during the 19th century by Lord Rayleigh and by Vincenz Strouhal of Czechoslovakia demonstrated that the frequency at which a cylinder sheds vortexes varies in proportion to the velocity of the wind.

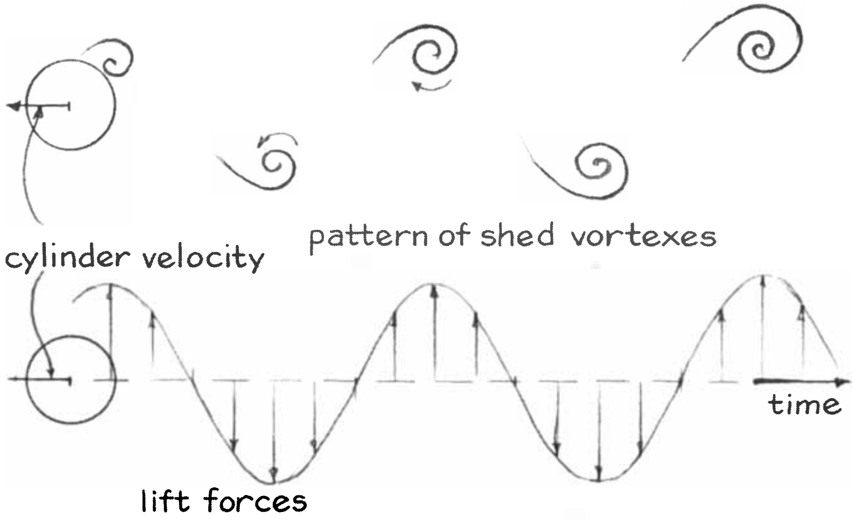
"My tests were made with cylinders of two diameters. The cylinders were supported in the wind tunnel by a hinged fixture that allowed the model to move only in the vertical plane, at right angles to the axis of the cylinder and to the direction of the airstream. The vertical motion of the model was limited by a set of helical springs. The tension of the springs determined the model's natural period of vibration. The natural period could be altered by substituting springs of different stiffness. The velocity of the airstream could be controlled by regulating the speed of the fan that sucked air through the wind tunnel.

"The alternating forces generated by the vortexes are relatively weak in a small wind tunnel. The model cylinders for my experiments were accordingly constructed of light material. Care was also taken to minimize the damping effect of the hinged fixture to which the models were mounted.

"The wind tunnel consists of four demountable elements [*see bottom illustration on this page*]. The rectangular entrance section curves inward, roughly in the form of an exponential horn. It is made of poster board mounted on a wood frame at the upstream end of the test section. The frame maintains the rectangular shape of the assembly. The frame can be detached from the test section for transportation.

"The open end of the entrance section is stiffened by strips of corrugated cardboard glued to the poster board. The corners of the board were similarly glued together. All joints were sealed with masking tape. The structure is surprisingly stiff and strong.

"The test section, a straight rectangular tube, is made of plywood. One exterior side supports a rectangular frame to which the fixture that supports the models is hinged [*see illustration on next page*]. The motion of the fixture is limited in the vertical plane by helical springs. Motion in the horizontal plane is prevented by a cord between the fixture and the frame on the upstream side.



Lift forces generated by the vortexes

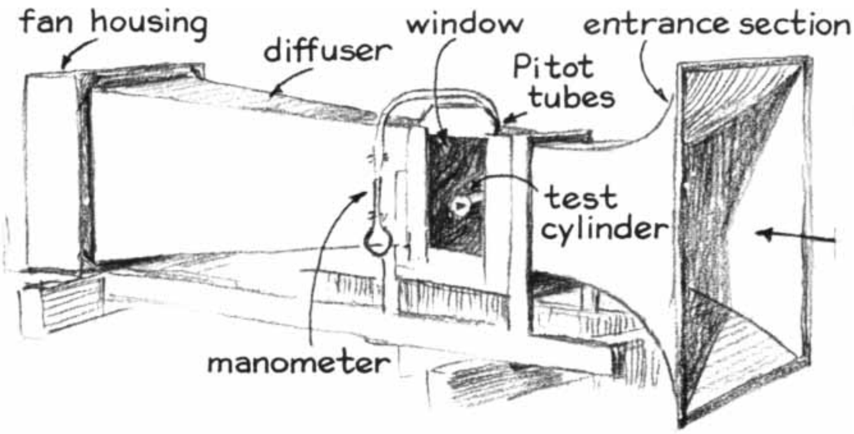
The cord is maintained in tension by a helical spring on the downstream side.

"Air pressure in the test section is measured by a manometer, a transparent tube in the form of a *U* that is partly filled with alcohol. The two arms of the manometer are connected to glass tubes with a bore of approximately six millimeters. One tube, which measures static air pressure, terminates inside the test section about four inches above the model and two inches upstream from it. The end of the tube is flush with the inner wall of the test section. The axis of the tube makes a right angle with respect to the direction of the airstream.

"The other arm of the manometer connects to a second glass tube of the same size. This tube extends through the wall of the test section approximately one inch from the static tube, but the glass is bent at a right angle and installed with the open end facing squarely into the wind. The tube measures total pressure. From the difference between the static pressure and the total pressure it is possible to calculate the velocity of the airstream, as I shall explain.

"The downstream end of the test section connects to the diffuser section, a trapezoidal tube that allows the air to expand gradually as it approaches the fan. The diffuser section creates maximum flow velocity within the test section. Maximum velocity is achieved by maintaining laminar flow between the test section and the fan, which should be larger in area than the test section.

"My first diffuser failed. Its sharp taper allowed the air to expand so quickly that separation occurred: the airstream pulled away from the walls, resulting in loss of energy and relatively low velocity. I replaced the section with a successful diffuser that tapered at an angle of seven degrees. The character of the flow was observed by attaching tufts of wool yarn to the end of a wire



Bonnie Jean Luessen's wind tunnel

rod and supporting them in the airstream.

"A single 21-inch floor fan, which is coupled to the downstream end of the diffuser, can suck air through the test section at a maximum velocity of 32.6 feet per second. The velocity can be increased to 40 feet per second by using two fans back to back. The speed of the fans can be adjusted within reasonable limits by a variable transformer. The velocity is calculated from manometer measurements by using Bernoulli's law: $p_s + \frac{1}{2}\rho V^2 = p_t$, in which p_s is the static pressure in the test section, ρ is the density of air, V is the velocity and p_t is the total pressure in the test section. The manometer indicates the difference in the pressure of p_t and p_s in inches of alcohol: $p_t - p_s = Dh$, where D is the density of alcohol and h is the difference in the height of the alcohol in the two arms of the manometer. Combining the equations and solving them for velocity gives $V = (2hD/\rho)^{1/2}$.

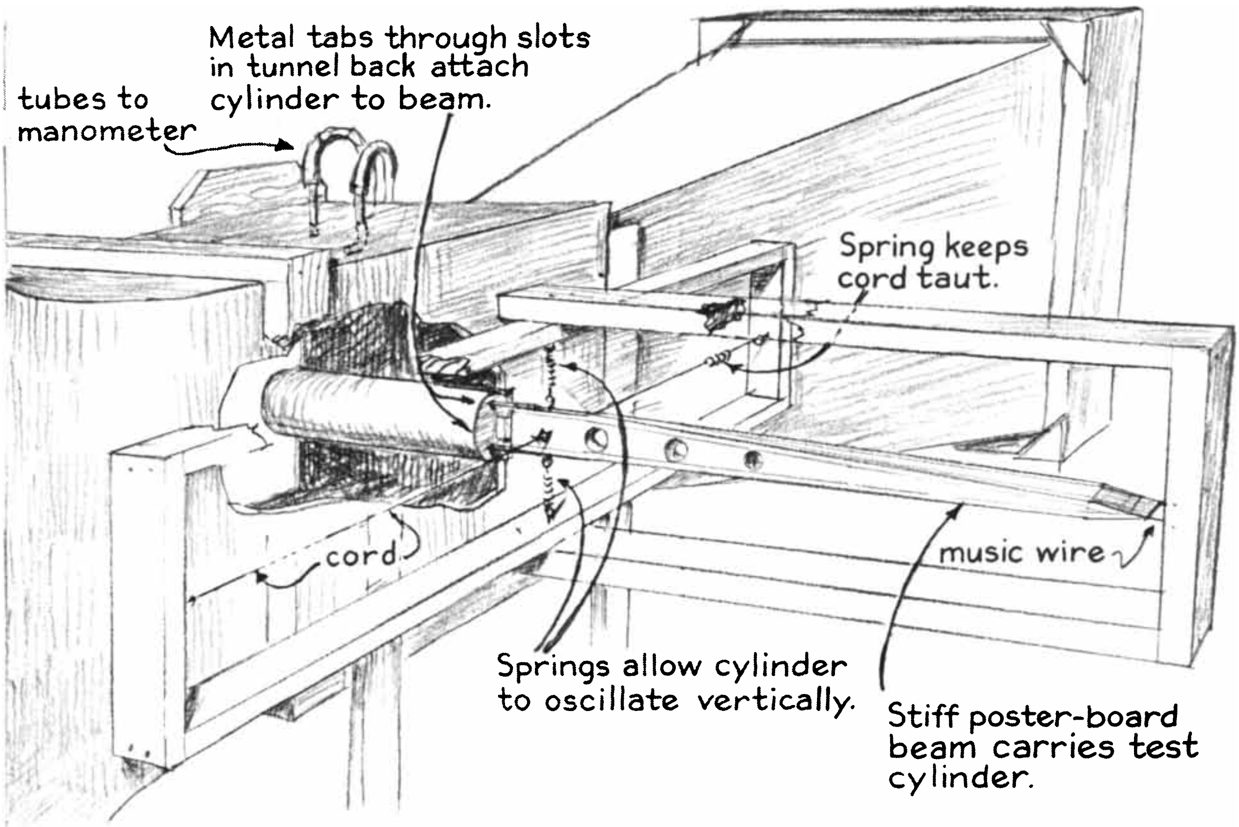
"I wanted to express velocity in feet per second. The density of alcohol in corresponding British units is 49 (pounds per cubic foot). The density of air is .002378 (slug per cubic foot). The difference in the height of the alcohol in the two arms of the manometer was measured in inches. The measurements were converted to feet (divided by 12) for use in the equation. For example, a manometer reading of .21 inch (.21/12 = .0175 foot) indicates a wind velocity in the tunnel of $(2 \times .0175 \times 49/.002378)^{1/2}$ = 26.9 feet per second.

"Since the forces developed by the vortexes are not large, the mass of the test cylinders was minimized by making the models of balsa wood. The fixture that supports the models in the test section was built of soft pine and poster board. Friction was minimized by coupling the distant end of the fixture to the frame with a spring hinge.

"Ideally the motion of the model should be pure translation. I approximated the ideal motion by making the fixture twice as long as the models. The models are attached to the fixture by a pair of metal strips that enter the test section through half-inch holes in the wall. The static pressure at this point in the test section is somewhat below atmospheric pressure. Some air doubtless enters the tunnel through the holes. I did not want to hamper the free movement of the models by sealing the openings with flexible membranes. Subsequent experiment demonstrated that the leakage was not sufficient to interfere with the vortexes.

"The free end of the model, which is visible through a window in the opposite wall of the test section, carries a triangle that functions as a vibrometer for measuring the amplitude at which the model vibrates. The altitude of the triangle is approximately one inch. The base measures .3 inch. The triangle is divided into three equal parts [*see top illustration on opposite page*]. When the model vibrates in the vertical plane, persistence of vision causes the eye to perceive two relatively light overlapping triangles plus a smaller dark triangle in the region where the two light triangles overlap. The altitude of the small dark triangle varies inversely with the amplitude at which the model vibrates. It is possible to estimate the altitude of the dark triangle to within 1/15 inch and therefore to estimate the amplitude of vibration to within .02 inch.

"The Langley Research Center of the National Aeronautics and Space Administration made a similar study of von Kármán's vortex street in 1966 as part of the Saturn rocket project. These tests were made with a cylinder three feet in



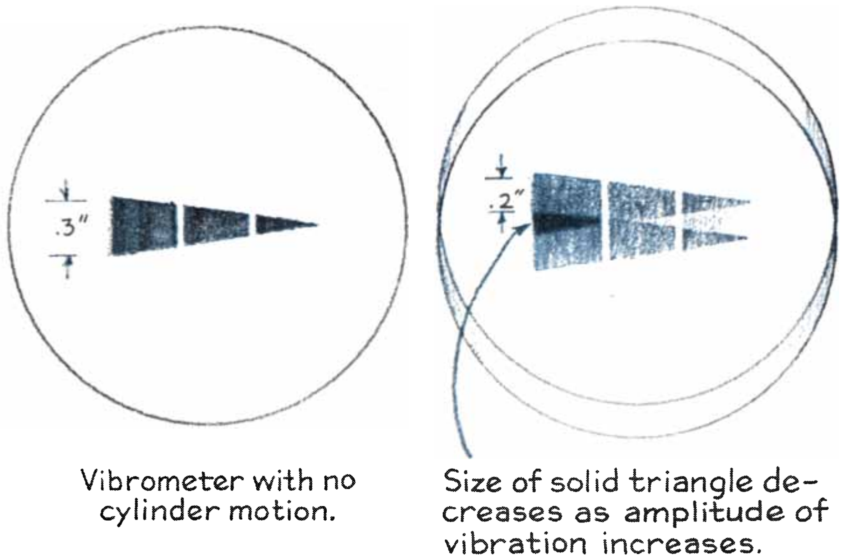
Details of the wind tunnel

diameter in a 16-foot wind tunnel. Since my tunnel could accept a model eight inches long, I scaled the cylinder in proportion, making one of my models 2.25 inches in diameter. A second model was made with a diameter of 1.75 inches for observing the influence of reduced diameter on the performance of the apparatus.

"At approximately what frequency should these cylinders vibrate in my tunnel? Strouhal demonstrated that the frequency with which a cylinder sheds vortexes varies directly with the velocity of the airstream and inversely with the diameter of the cylinder. The relation is described in terms of a constant, which is known as the Strouhal number. It is the product of the vortex frequency times the diameter of the cylinder divided by the velocity of the airstream ($S = fd/V$, in which S is the Strouhal number, f the frequency, d the diameter of the cylinder in inches and V the velocity of the airstream in inches per second). The value of the constant is .2. The frequency at which the cylinder should vibrate is therefore equal to $.2 \times V/d$.

"With the maximum velocity of my wind tunnel, running on one fan, being about 32 feet per second, if one assumes a flow of 25 feet (300 inches) per second a model 2.25 inches in diameter should oscillate at the rate of $.2 \times 300/2.25 = 26.6$ vibrations per second. This frequency seemed reasonable for a small cylinder of balsa wood. The problem of suspending the model with springs of the stiffness required for a natural period of about 26 vibrations per second was solved experimentally. I merely suspended the model with a set of springs, plucked it with my finger, measured the rate of the resulting vibration with a stroboscope and then substituted springs of greater or lesser stiffness as required to generate the desired frequency of about 26 vibrations per second.

"A model so tuned might tend to oscillate in response to mechanical vibrations developed by the fan or some other source. On the other hand, if the amplitude of vibration approaches maximum as the velocity of the airstream approaches the critical value predicted by the equation and thereafter declines as the velocity increases, one could conclude that the oscillations are induced by von Kármán's vortex street even though the presence of the vortexes is not confirmed by a different technique, such as injecting smoke into the wake of the cylinder. In order to prove the existence of the vortexes without resort



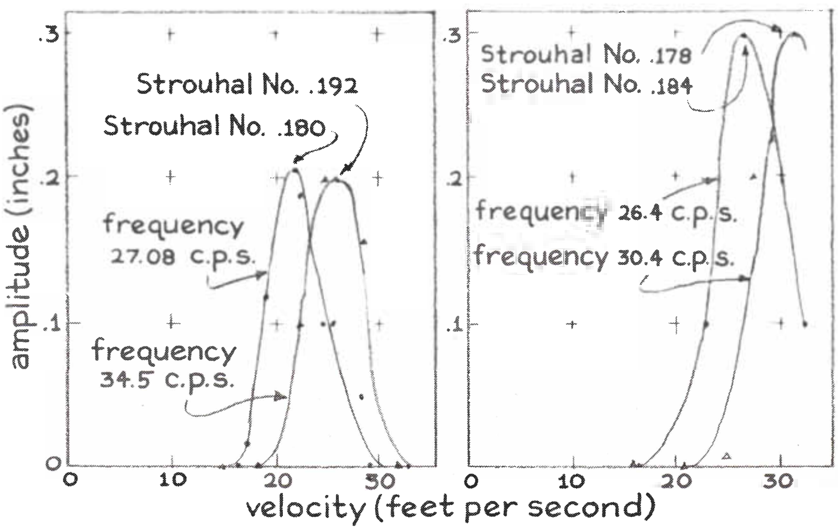
Arrangement of the vibrometer

Cylinder diameter = 1.75 inches
Model frequency = 27.08 cycles per second
Critical velocity = 19.63 feet per second (for S = .2)

Manometer reading (inches)	Amplitude response of model (inches)	Airflow velocity V (ft/sec)	Calculated Strouhal No. S (dimensionless)
.08	0	16.6	.238
.09	.02	17.6	.225
.11	.12	19.5	.215
.14	.21*	22.0*	.180*
.15	.18	22.7	.175
.18	.10	24.8	.159
.19	.10	25.6	.1545
.22	.05	28.1	.141
.24	0	28.8	.137
.30	0	32.6	.121

*denotes critical measured quantities

Data from wind-tunnel tests



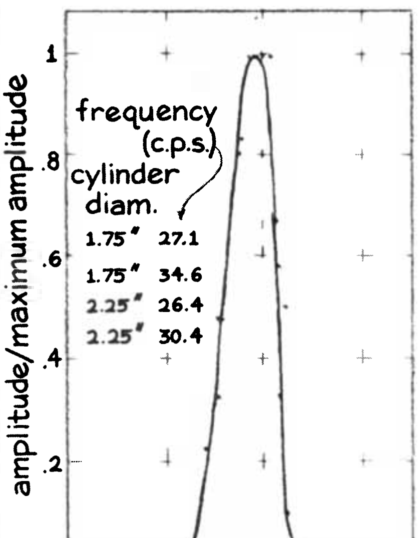
Amplitude in relation to velocity

to such alternative techniques I made up four model configurations.

"Two cylinders of 2.25-inch diameter were tuned to resonate respectively at 26.4 and 30.4 cycles per second. Two 1.75-inch cylinders were tuned to 27 and 34.5 cycles per second. The amplitude at which each cylinder vibrated was observed at 10 discrete velocities as the air speed was increased in increments through the range from approximately 16 to 32 feet per second. At each velocity I tabulated the air pressure as indicated by the manometer and the amplitude of vibration as indicated by the vibrometer.

"From these data I calculated the corresponding air speed and Strouhal number. The accompanying table [*middle of preceding page*] lists a typical set of data. The results of the four tests were displayed as graphs made by plotting amplitude versus velocity [*see bottom illustration on preceding page*]. During each test the amplitude of vibration increased to maximum and then decreased as the velocity of the airstream increased from minimum to maximum. The frequency of vibration remained constant, as expected at the rate to which the models were tuned. The measured value of the Strouhal number at which maximum amplitude was observed ranged from .178 to .192. The Strouhal numbers averaged within 8.3 percent of the theoretical value of .2.

"Finally, the data of all four tests were combined in a single graph [*see illustration below*]. The amplitude at which each model vibrated at each increment of air speed was divided by the maximum amplitude of that model. The quotients were plotted against Strouhal numbers ranging from .12 to .24. The resulting graph peaks at a Strouhal number of .19, within 5 percent of the theoretical value, indicating that the vibrations were indeed generated by the aeolian effect. Moreover, the results have assured me that I can use the small wind tunnel with some confidence for doing other aerodynamic experiments, the outcome of which may not be so easily verified by theory."



Combined results

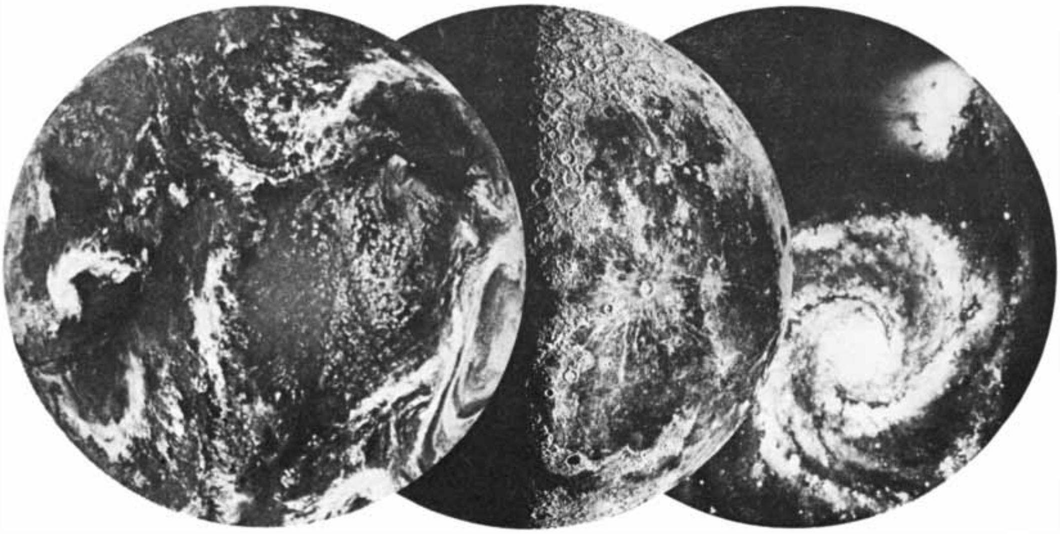
John Livingood of Hinsdale, Ill., submits a simple design for a Cartesian diver together with an amusing elaboration on the experiment. "The diver," he writes, "is made from a spherical Christmas-tree ornament about half an inch in diameter. The buoyancy of the diver is adjusted by attaching a length of wire to the metal loop at the neck of the ornament and cutting off bits of the wire until the bulb barely floats neck down in water. The remaining wire can be coiled in a convenient bundle.

"Fill a bottle with water to within an inch or so of the top and insert the diver. Place a snug-fitting stopper in the neck of the bottle. Press the stopper gently. The increased air pressure forces water up into the ornament. The diver sinks. Releasing the pressure on the cork causes the diver to rise. If the bottle is filled with water completely to the stopper, pressure applied to the stopper will be transmitted directly to the air in the diver. In this state the device is inconveniently sensitive and difficult to control.

"It is easy to demonstrate that the diver is unstable at any depth. If the diver starts to sink at any position of the stopper, the pressure of the water increases, more water enters the bulb and the bulb loses buoyancy and sinks deeper. Conversely, when the diver starts to rise, the pressure of water at the neck of the bulb decreases, the trapped air expands, water is ejected and the rate of ascent increases.

"The diver can be stabilized, however, by altering the density of the fluid. Fill the bottle to about half of its capacity with fresh water and insert a funnel with a stem long enough to reach the bottom [*see illustration on page 112*]. (A long-stemmed funnel can be improvised by taping a soda straw to the stem of a conventional funnel.)

"In a separate container make up a brine solution by dissolving a few table-



spoons of salt in fresh water. Chill the brine with a cube of ice. By means of the funnel slowly add chilled brine to the bottom of the bottle until the upper surface of the fresh water floats to within an inch or so of the top of the bottle. By looking through the bottle toward a source of light you can see the interface between the brine and the fresh water as a result of the effects of optical refraction.

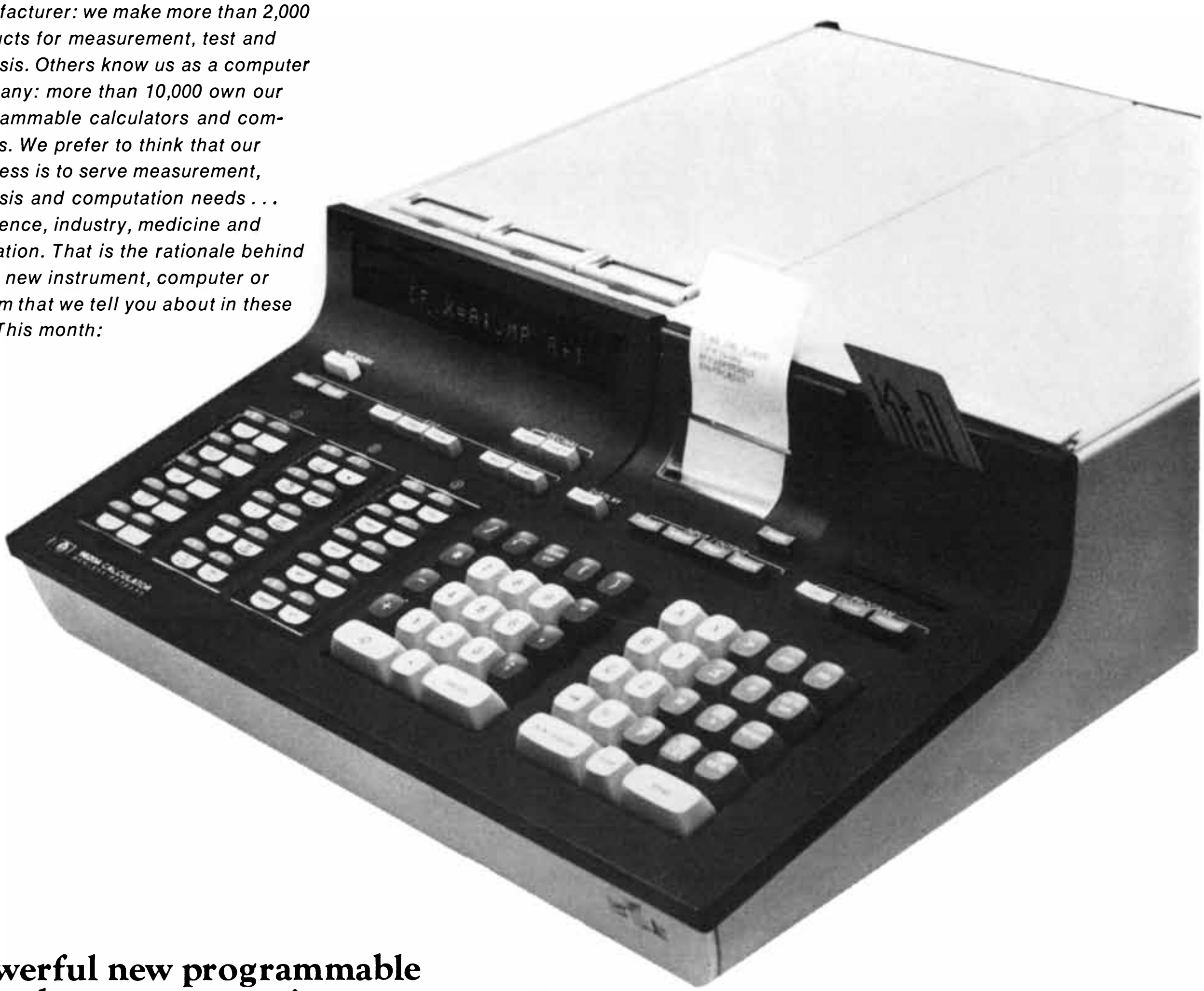
"Make a stirring rod by taping a dime across the end of a soda straw. Insert the rod in the water and gently lower the dime to the level of the interface. Stroke the rod slowly up and down about a dozen times, beginning with quarter-inch excursions and ending with two-inch excursions. Remove the rod. This action, when it is carefully performed, creates a zone of increasing density that is approximately five inches thick.

"Observe that the diver is now stable at any depth within the five-inch zone at which you place it by manipulating the stopper. When a slight pressure is exerted on the cork, the diver first sinks and then rises and oscillates about a new position of equilibrium. Increased pressure forces solution into the diver, causing it to lose buoyancy and sink, as in the first experiment. This time, however, the diver sinks into solution of increasing density, where it gains disproportionate buoyancy. The diver rises. After passing through the zone of equilibrium it again loses fluid, but the surrounding solution is less dense, so that the motion again reverses. After a few oscillations of diminishing amplitude the diver comes to rest at the new level of equilibrium. The density gradient will usually persist in the solution for a week or more before it is destroyed by diffusion."

John Livingood's Cartesian diver

Some things are changing for the better.

Many people know us as an instrument manufacturer: we make more than 2,000 products for measurement, test and analysis. Others know us as a computer company: more than 10,000 own our programmable calculators and computers. We prefer to think that our business is to serve measurement, analysis and computation needs . . . in science, industry, medicine and education. That is the rationale behind every new instrument, computer or system that we tell you about in these ads. This month:



Powerful new programmable calculator converses in simple algebraic language.

Hardly a month goes by that doesn't signal the introduction of a new calculator with more power, more memory, more output flexibility. But are things really changing for the better when the improvements so complicate the use of a calculator that one must practically become a computer programmer before he can harness its power?

This is precisely where HP's new Model 20 makes its most significant contribution. It is easier to use, by far, than any other calculator. Its language is the most simple: algebra, the kind you learned in high school. Not only does it 'understand' algebra, it also 'speaks' it, using the same numbers, letters and symbols that you do.

You enter an equation just as you write it on paper including implied multiplication and nested parentheses. The Model 20 displays your entry for verification the same way you wrote it. For example:

(-B+r(BB-4AC))/2A

It tells you what to do next (in the program mode):

ENTER A

Then you enter the values on the keyboard (for A, B and C in this example) and press a key to execute. The calculator immediately displays and prints the solution to ten significant digits, along with its English label if you desire:

REAL ROOTS

If you make a mistake, the Model 20 tells you so and identifies precisely what the error is . . . lets you correct it without redoing the entire line, let alone the entire program . . . and automatically adjusts program storage to occupy the least possible memory.

In addition to its conversational ability, the Model 20 changes things for the better not only through more power and more memory but through a hardworking line of Series 9800 Peripherals: X-Y Plotter, Typewriter and Card Reader, to name a few. Model 20 costs $5,475.

Instrumentation quality tape recording at a bargain price.

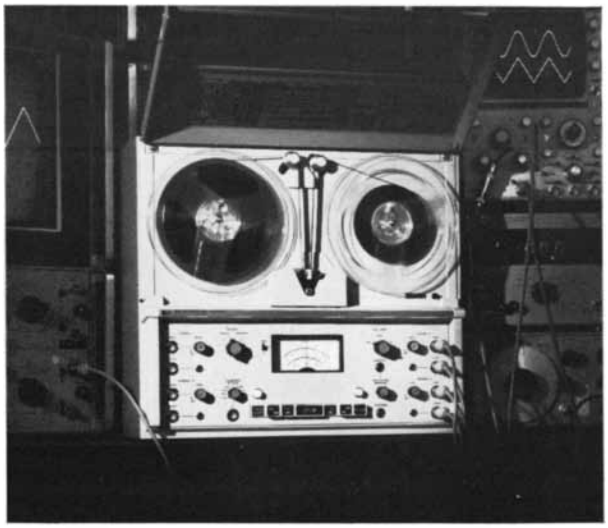
Most scientists would use portable instrumentation tape recorders for analog recording if only they performed as well as the big expensive laboratory machines. Unfortunately, their small size usually meant reduced performance.

Then came the HP 3960. Truly portable in size (50 pounds) and low in price ($4,270 for a fully-equipped four-channel instrument), the 3960 actually outperforms most laboratory machines costing five times more.

If this sounds too good to be true, listen to some of the 3960's capabilities. At 15/16 ips, its FM signal-to-noise ratio of better than 200:1 lets you play back signals that would be buried in noise (ECG's for example) on many lab machines as well as on any other portable.

The 3960 lets you mix and interchange four FM direct record/reproduce channels at will. You have a choice of three electrically-switched speeds, for a time-base expansion of 16:1 or 10:1 . . . without signal degradation. Tape drive is bidirectional so that you don't have to rewind either to continue recording or to play back.

Built-in facilities let you calibrate the 3960's FM electronics without external equipment. And an integral peak-reading meter lets you optimize record level without using a scope. Options include a 5 to 30 foot loop adaptor, an interrupting voice channel, and an inverter for 12 or 28 VDC . . . all integrally mounted.

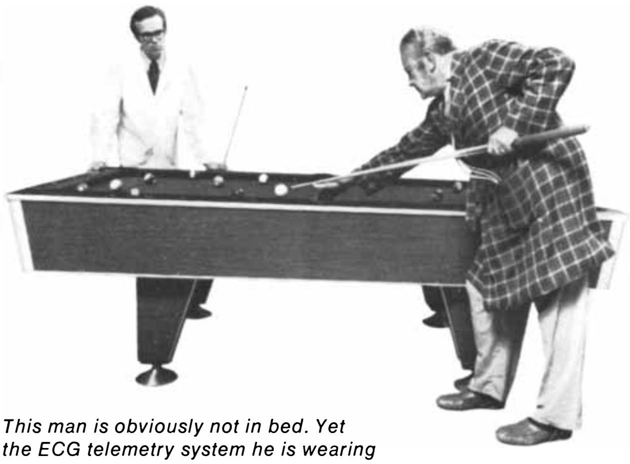


Write for Application Note 89, a tape recording handbook useful to scientists interested in tape recording techniques for vibration and test analysis, research and clinical medicine, acoustics, oceanography and other environmentally difficult research projects.

For complete information, write Hewlett-Packard, 1506 Page Mill Road, Palo Alto, California 94304. In Europe: 1217 Meyrin-Geneva, Switzerland; Japan: YHP, 1-59-1, Yoyogi, Shibuya-Ku, Tokyo, 151.

00212

HEWLETT hp PACKARD



This man is obviously not in bed. Yet the ECG telemetry system he is wearing enables nurses at a central monitoring station to keep close watch on his heart action.

Freedom with protection for the post-coronary patient.

Once the coronary patient is released from the intensive care unit, his recovery can often be aided by freedom to move about and mild exercise . . . provided his ECG can be continuously monitored.

With the new HP ECG Telemetry System, the post-coronary patient can be ambulatory. Wherever he goes, his heart action is transmitted to a receiver at the nursing station where it can be continuously observed. The transmitter is small enough to be carried comfortably in a bathrobe pocket, has a strong enough signal to reach the nursing station from 200 feet even through several masonry walls, and is rugged enough to operate reliably even if dropped.

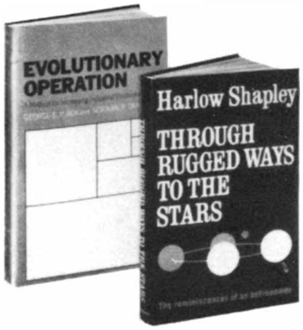
At the nursing station, the patient's ECG signal is monitored by a receiver that operates automatically, never requires tuning and accepts only valid signals, minimizing artifacts from patient motion. An automatic warning light alerts the nurse of such inoperative conditions as: patient out of range, dislodged electrode, low battery power. It is completely compatible with HP patient monitoring systems. Because it doesn't require new wiring, the ECG Telemetry System is easily introduced into existing facilities. Price is $1,800 for each patient unit. Write for our new illustrated brochure.

The most exciting value ever to help you build your own science library!

(Retail prices light, member prices bold.)

79490. SOURCE BOOK IN ASTRONOMY. 1900-1950. *Edited by Harlow Shapley.* $11.95/**$7.50**

84270. THROUGH RUGGED WAYS TO THE STARS/OPPENHEIMER. Harlow Shapley reminisces. *Plus* a new look at J. Robert Oppenheimer, by I. I. Rabi, Robert Serber, Glenn T. Seaborg, others. *The two books count as one.* $12.90/**$8.50**



46590. EVOLUTIONARY OPERATION. *George E. P. Box and Norman R. Draper* use statistics and the basic concepts of natural selection in this coordinated approach to process improvement. $11.00/**$8.50**

46580. THE EVOLUTION OF MAN AND SOCIETY. *C. D. Darlington.* $12.95/**$9.75**

70360. AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS, Volume II, Second Edition. *William Feller.* $15.95/**$10.95**

84240. THE PHYSICS OF MASS, LENGTH AND TIME/THE PHYSICS OF VIBRATIONS AND WAVES/ELECTRICITY AND MATTER. *Norman Feather, F.R.S. The three volumes count as one book.* $19.85/**$13.25**

62100. MEN AND DINOSAURS: The Search in Field and Laboratory. *Edwin H. Colbert.* $8.95/**$6.95**

74150. THE RISE AND FALL OF T. D. LYSENKO. *Zhores A. Medvedev. Translated by I. Michael Lerner.* $10.00/**$7.75**

63340. MUSIC BY COMPUTERS. *Edited by Heinz von Foerster and James W. Beauchamp.* Includes four 33⅓ rpm records. $14.95/**$9.95**

56350. INTRODUCTION TO GEOMETRY, Revised Edition. *H. S. M. Coxeter.* Urbane, unconventional, witty—a modern classic. $11.95/**$7.95**

Take any 5 books FREE (values to $83.45)

with your first selection (at low discount price) as a member of

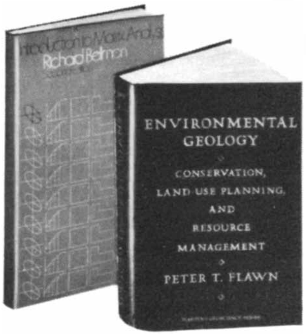
The Library of Science

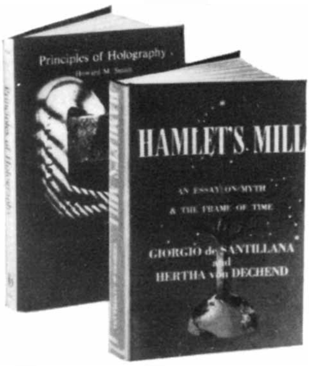
and you need buy only 3 more books at member discount prices in the next 12 months!

61330. THE MATHEMATICAL SCIENCES: A Collection of Essays. By America's most distinguished mathematicians. $8.95/**$6.95**

56540. INTRODUCTION TO MATRIX ANALYSIS, Second Edition. *Richard Bellman.* $14.75/**$10.50**

45490. ENVIRONMENTAL GEOLOGY. *Peter T. Flawn.* Relationship between man and natural earth processes. $13.95/**$9.95**





70070. PRINCIPLES OF HOLOGRAPHY. *Howard M. Smith.* The first unified and complete treatment of this newest of sciences. $9.95/**$7.50**

52180. HAMLET'S MILL: An Essay on Myth and the Frame of Time. *Giorgio de Santillana and Hertha von Dechend.* A brilliant speculative inquiry into the origins of scientific thought in the preliterate world. $10.00/**$7.75**

MEMBERSHIP APPLICATION
The Library of Science, Riverside, N.J. 08075 2-647

Upon acceptance of this order, please enroll me as a member and send the books I have indicated. Bill me only for my first selection at member's price, plus postage and handling. If not delighted, I will return all books within ten days and this membership will be cancelled.

As a member, I need accept only three more selections during the next 12 months at reduced member prices, plus postage and handling. Savings range up to 30% and occasionally even more. I understand that I will receive free advance Reviews which fully describe each month's Main Selection and Alternates. If I wish the Main Selection, I need do nothing and it will be sent automatically. If I prefer an Alternate—or no book at all—I need only return the convenient reply card you send me by the date specified. I understand that I may choose a bonus book for every four selections purchased. (Introductory offer counts as first selection toward a bonus book.) Send no money. Members are billed when books arrive.

My 5 FREE books are: (write in numbers)

____ ____ ____ ____ ____

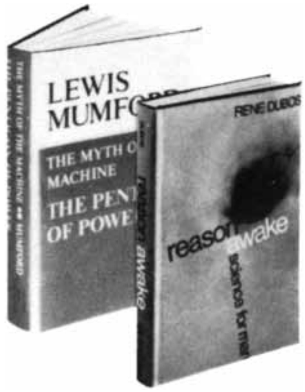
My first selection is (write in number) ________

Name ________

Address ________

City ________ State ________ Zip ________

(Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.)



67650. THE PENTAGON OF POWER. *Lewis Mumford.* Examining the fundamental difference between the good life and the "goods life" of modern culture. $12.95/**$8.50**

73490. REASON AWAKE: Science for Man. *René Dubos.* Penetrating analysis by a Pulitzer Prize-winning humanist-scientist. $6.95/**$5.95**

56320. INTRODUCTION TO CALCULUS AND ANALYSIS. *Richard Courant and Fritz John.* $11.95/**$8.50**

68730. PHYSICS OF THE EARTH. *Frank D. Stacey.* $11.95/**$7.95**

61320. MATHEMATICAL SNAPSHOTS, Third Revised and Enlarged American Edition/SECOND MISCELLANY OF PUZZLES. Varied feast of choice math gems—*Plus* devilish, original puzzles. *The two books count as one.* $12.45/**$8.50**

79470. A SOURCE BOOK IN MATHEMATICS. *D. J. Struik.* Big, 472-page volume of original selections. $11.95/**$8.25**

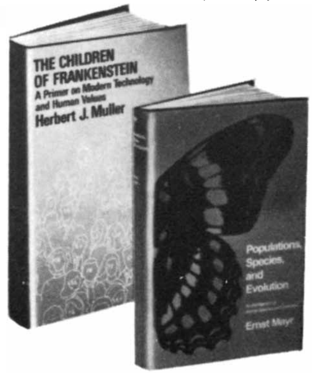
79510. A SOURCE BOOK IN CHEMISTRY, 1400-1900. *Edited by Henry M. Leicester and Herbert S. Klickstein.* $11.95/**$8.50**

39770. COMPLETE RUSSIAN COURSE FOR SCIENTISTS/RUSSIAN-ENGLISH TRANSLATORS DICTIONARY. *The two books count as one.* $17.95/**$10.95**

79520. SOURCE BOOK IN GEOLOGY 1900-1950. *Edited by Kirtley Mather.* $12.50/**$8.95**

69930. POPULATIONS, SPECIES AND EVOLUTION. *Ernst Mayr.* $10.00/**$7.75**

37880. THE CHILDREN OF FRANKENSTEIN: A Primer on Modern Technology and Human Values. *Herbert J. Muller.* $10.00/**$7.75**



BOOKS

Scratches on a Paleolithic eagle bone may be the first calendrical notation

by Philip Morrison

THE ROOTS OF CIVILIZATION: THE COGNITIVE BEGINNINGS OF MAN'S FIRST ART, SYMBOL AND NOTATION, by Alexander Marshack. McGraw-Hill Book Company ($17.50). A single photograph fills the big marginless page with dark dagger forms on a textured ivory ground. This large, handsome volume of many photographs and drawings does not, however, celebrate a school of modern painting. The photograph is a magnified portion of a hollow eagle bone, held in Cabinet No. 1 at the Musée des Antiquités Nationales near Paris. The eagle flew some 13,000 to 15,000 years ago ("late middle Magdalenian"). The bird was killed; one hard bone was cut and scratched with cunningly worked flint edges and points. Across the bone now march three lines of specific markings, a few of which, very much blown up, form the picture. That four-and-a-half-inch piece of bone, photographed at many different scales, drawn and redrawn, the marks counted, grouped and plotted, forms a central pillar of this remarkable monograph.

The eagle bone bears engraved marks that could not plausibly have been made by a careful artisan of the Magdalenian working for a few hours simply to decorate the bone. First of all the engraved marks can be identified under the microscope not with one tool point, or even a few, but with "13 to 14 changes of point on this single fragment." All the small lines of one series, say, were made with a single point. Then another point was used to make one or two dozen small strokes. It seems likely that different points were used to make the marks from month to month over a considerable period of time.

Second, there is a sister bone: another eagle bone from the same rock-shelter, from the same level, perhaps from the same hand. (The discovery of the second bone was itself melodramatic. The clue was a picture of it, similar and yet not identical, in a book published in 1907. The collection from which it came had been closed to all examination "by will of Edouard Piette, implemented by the Countess de Cugnac," Piette's granddaughter. "No one could approach the materials" because certain pieces had been damaged in an earlier effort at photography. It took a year and a half to obtain permission "merely to look through the glass of the cabinets. The piece was missing." Marshack found it accidentally in a tiny collection in Poitiers, where it had been sent on loan 40 years earlier.) That second eagle bone bears three analogous sets of markings, again done with multiple points. "Clearly, what we have is notation...." It seems unlikely that art or whim could produce two such neatly kept "records." Magic could be invoked; what is not a magical object? Such magic is still notation, still a deliberate string of symbols.

The third argument is the tentative identification, perhaps less secure than the looming fact of so ancient an abstract and arithmetical symbolism, of these careful stroke sequences with the successive days on which the moon can be seen. Indeed, the strokes show groupings and changes in sequence whose length is now seven, now 15, now 29. Marshack presents a lunar model for all these numbers, relating the changes to the phases of the moon, associating the gaps with the variable periods before the new moon when the moon is invisible and adding the strokes to reach just 12 lunar months on one eagle bone and 13½ on the other.

Thirteen and a half months? That touches the weak part of the impressive argument. Some dozen artifacts of the mobiliary art of the upper Paleolithic and Mesolithic are analyzed in numerical detail. Every small stroke is counted; long and apparently marginal lines are plausibly held to be outside the reckoning. The counts usually fit the lunar model quite well, but that model is itself randomly varying, weather-dependent and of optional length, and is built on incommensurable rhythms that demand gaps and sightings. Without real statistical studies, which would be premature, we cannot rely on the lunar phasing, as the author quite openly admits. The suggestiveness of his interpretation lies just short of conviction. More work! We wait patiently for the sequel, for other microscopic studies, for the analogies from the Nicobar Islands and from American Indian calendar sticks; all are promised.

One ought to mention more elaborate calendrical records that appear in the book, done with tiny images of plants budding in the spring, seasonal groupings of animals and the like. By the Mesolithic an annual calendar is almost detectable. The Venus figures, the *bâtons de commandement* and other familiar classics of the cave are also studied and sifted, with some yield but nothing like a nugget of certainty. Marshack is a gifted amateur in a domain where the memory of amateur success is as green as the epic of Troy. He began as a writer musing on the prehistory of knowledge of the moon. He initiated his hypothesis on the basis of a photograph in *Scientific American*. He first went to the dusty basements of the French museums with so little support that he had to use a tiny Japanese toy microscope. Now he is a pro, with stipend and rank, and several monographs in French behind him. French and American professionals were early, even generous, in accepting him. A general reader comes easily to a verdict: There *is* abstract symbolism in the Old Stone Age. It may well be—but it is not clearly—lunar counts made night after night. Notation is as old as the caves, and so, perhaps, is astronomy.

The book is beautifully, almost lavishly, presented. For that we can thank the publisher and the editors. But the overblown title and the entire *ambiance* of the book suggest a somewhat unkind effort to place the volume in the hands of many readers anxious to see art and to read drama. Although the personal adventures of the author are winning enough, much of the text is given to counts and trials that are more tedious than the narrative of the usual archaeological best sellers. Although we owe the striking presentation of Marshack's book to this marketing stratagem, one might complain a little on behalf of less earnest readers. Those who have the patience to follow Marshack's argument will find his book a pioneer, a beauty and a new door ajar.

SYMMETRY: A STEREOSCOPIC GUIDE FOR CHEMISTS, by Ivan Bernal, Walter C. Hamilton and John S. Ricci. W. H. Freeman and Company ($9.50). MOLECULAR REALITY: A PERSPECTIVE ON THE SCIENTIFIC WORK OF JEAN PERRIN, by Mary Jo Nye. American Elsevier Inc. ($15.50). The core of *Symmetry,* a handsome, slender introduction to the subject for students of chemistry, is a kind of atlas. On each of some 40 pairs of facing pages there appear a brief comment on one after another of the three-dimensional point groups, a computer-produced stereoscopic pair illustrating the group by a suitable array of a single basic unit, a paragraph on the chemical interest of a particular molecular species that exhibits the symmetry in question, and a stick-and-ball diagram of the molecule, again in a computer-drawn stereoscopic pair. The open pages, the crisp, contrasty typographic presentation and the splendid spatial illusion provided by the viewer in each copy of the book make looking at these molecules a delight.

The book includes a brief but sensible introduction to the concepts of symmetry and of groups of symmetry operations. It presents in addition to the molecular examples the 10 two-dimensional-point groups that are compatible with a plane crystal lattice, and a sample of four of the 17 possible two-dimensional crystal groups, realized in the footprints of some mathematical Man Friday. The use of a human footprint to serve as a unit is characteristic of the lighthearted text, which displays three faces of Edgar Allan Poe, one as he was photographed and one each reassembled out of the left and right sides of his face. (His asymmetric mustache plainly suggests that Poe cultivated the departures from bilateralism that human beings universally possess.) The annotated bibliography of course includes the work of the late M. C. Escher, and the crystallographic analysis of his plane drawings by C. H. MacGillavry. In so visual a subject the authors do not hesitate to cite an animated film (the prizewinning *Symmetry,* directed by Philip Stapp, "a sheer delight for all ages").

Some of the molecules illustrated are worth recommending to viewers. Cubane, C_8H_8, a "beautifully symmetric hydrocarbon," is a cube with eight carbon corners; on each corner is a hydrogen atom floating out along the cube's threefold rotation axes. The chemical tyro misses comment on the properties of the stuff; what, for instance, does the elegant cubane smell like? A molecule with the formidable name of hexapyridineiron (II) tetrairontridecarbonylate, if it were executed in welding rod or wrought iron, would sell easily, name and all, in the art galleries.

The authors mention only here and there the evidence that lies behind each stereoscopic drawing. Much of the work rests, of course, on X-ray diffraction in crystals, which is a kind of photography in a real enough transformed space. But a good deal rests, as it has for a century, on that abstract logic which synthesizes a model in space out of a long series of chemical syntheses, a cunning inference from products that formed and did not form. These molecule models of the chemist—three-dimensional but stationary on the page, neither spinning nor drifting, represented not by realistic charge clouds bulkily filling space but by clean abstract linear bonds between neat atomic circles—are strongly idealized. Real they are today (they were always real to the chemist) but to the less imaginative they are only thumb-sized figments, mere logical structures not deemed worthy of the word "reality" by any critically empirical thinker.

Such was the frame of mind at the turn of the century. Chemists reveled in their synthetic powers and physicists watched the marvels of X ray and radium, electron and ion. But an articulate school, foreshadowed by Heinrich Hertz and fully formed by Wilhelm Ostwald in Germany, Ernst Mach in Austria and Pierre Duhem in France, would have none of the unseeable three-dimensional forms men called molecules and atoms. Those were powerful voices, and they retain some impact still. Men such as Max Planck wrote modestly of the kinetic hypotheses; Ludwig Boltzmann himself was the target of "vicious polemics" for his espousal of molecular reality. Boltzmann died a suicide.

That era is evoked in Professor Nye's detailed and interesting narrative of the life and work of Jean Perrin. It can be said that Perrin, mostly in the years from 1903 to 1913, gave decisive and graphic support to the reality of the kinetic, molecular world. Ironically enough, he used a physical model: visible colloidal particles wandering erratically in a molecular fluid. The molecules are quite invisible, but the dark field of the ultramicroscope displays, if not images, bright diffraction disks of particles only a small fraction of a light wavelength in diameter. For two decades men such as J. Thirion and Léon Gouy had singled out the Brownian motion as the token of molecular reality. These workers had no numbers. The velocities of the Brownian particles were hard to measure; the entire system seemed dauntingly complex.

Perrin, independently of Einstein's earlier publication in the famous year of 1905, saw that the "atmosphere" of colloidal grains would show an exponential decrease in density with altitude, like the barometer law in the idealized earth's atmosphere dear to every textbook. He worked with great care, making tens of thousands of observations by photography and by marks on projected images, flashing his light source to avoid heating effects and preparing his pigment particles by repeated centrifugation (with a yield of only one in a few thousand of the starting amount by mass but in uniform granules well under a micron in size). His experiments worked. One striking result can be seen in the figures reproduced here from the papers of 1909: the scale height of that atmosphere of yellow gamboge "molecules" is microns, not kilometers. Perrin went further: he also measured the Brownian rotation of large granules (12 microns in diameter), which he learned how to prepare.

By this time (1908) three powerful theorists had given in detail broadly concordant treatments of the Brownian motion. They were Einstein (who had the priority), Maryan Smoluchowski and Paul Langevin. Einstein wrote Perrin in 1909: "I had thought it impossible to investigate Brownian movement so precisely." There remained in 1909 a 15 percent spread between values of Avogadro's number gained from the Planck-Lorentz theory of the black-body radiation and Perrin's theory. Perrin was confident the gap would close; Einstein saw the test as being critical. By the end of 1912 the philosophically sophisticated Henri Poincaré was convinced: "Atoms are no longer a useful fiction; things seem to us in favor of saying that we see them since we know how to count them. ... The atom of the chemist is now a reality." And in his 1913 book *Les Atomes*, which received "an almost unanimous crescendo of applause," Perrin could list 13 ways of counting to one mole, which differed in their extremes only between six and 7.5 times 10^{23} atoms per mole. (We can specify that number today to one part in 10,000.)

Ostwald recanted early, Duhem and Mach never. The physicist son of Jean Perrin, Francis Perrin, himself studied under Henry Le Chatelier at the Sorbonne in the 1920's. The venerable Le Chatelier was still skeptical of atoms; he held that the helium reported in radioactive decay must have been present as impurities in the sample! But the mechanical model—"the engineer's universe"—has had no simple triumph. Space and time fused in the hands of Einstein. Probability reigns, and our quantized fields are far from the mechanisms that James Clerk Maxwell used to draw to explain electromagnetic fields. Discontinuity is real, whether or not the continuum of space and time remains so. Physics evolves in widening gyres, and when Jean Perrin died in New York in 1942, a fugitive from the rude invasion of his homeland, he left a legacy of permanent validity.

This is a historian's book; the index and the presentation of mathematics will be somewhat inadequate for scientist readers.

THINK TANKS, by Paul Dickson. Atheneum ($10). Out of the Douglas Aircraft Company by the Army Air Forces, RAND was born at Santa Monica in the months after World War II. The first charge to the organization (it is still the contract that pays half of the bills) was for " 'a program of study and research on the broad subject of intercontinental warfare other than surface. . . .' " A fanatically sectarian college without students, wary of critics, privy to ambitious air generals, RAND (for Research ANd Development) has spun out the theoretical base for the arms race, and has devised, or at least justified, most of the gambits of the Strategic Air Command, from refueling, hardening and fail-safe to how best to split the fallout between our old enemy Red China and our old friend Japan. It has often revealed real insight: as early as 1946 RAND saw the magic that would attend the first artificial satellite, and later RAND people foresaw and indeed developed much of the conceptual design of satellite reconnaissance. But its strategic policy innovations, whether they arose more from the sober calculations of Albert Wohlstetter or more from the apocalyptic world view of Secretary of Defense James V. Forrestal, have rationalized the arms race ever since 1951. RAND largely sees the nation's problem as lying in the grim fact of Russian nuclear weapons. It may well be, however, that RAND itself is more the problem than the solution.

Dickson, a young free-lance writer who has made his excellent journalist's report mainly out of visits and interviews and rather less from the documentary record, sums up RAND today: "It has by its own admission fallen behind in perceiving and acting to solve the major problems facing the nation. It is now trying to catch up, and without question a new RAND is developing.... Its once single-minded concern will soon become just one of many concerns about national security in the broadest, nonmilitary sense of the term. If RAND cannot make that conversion, then it knows it might as well go out of business." Perhaps it could start making typewriters, he adds, embroidering an old jape about RAND's name and public image. But can the B-52 ever change its spots?

Mother RAND has many children, some born before her. There are "at least 17,000 research entities" in the U.S. today, from the Hale Observatories to the dog-food-company laboratories. About 600 of them are think tanks. These are here defined as multidisciplinary R&D groups, closely coupled to industry, government or public purpose, whose output is not directed toward science and technology so much as toward analytical matter "relevant to people who make policy," from admirals to mere voters. About 75 think tanks are contract adjuncts to Government agencies, like RAND; another dozen or so live within the Government itself, like the Army's Institute for Land Combat; 200 are nonprofit enterprises, like the Stanford Research Institute; 300 more simply think for profit as *condottieri* of the brain. A handful of them, like the Institute for Policy Studies—the left end of the think-tank continuum—work and publish for a point of view they hold as the public good.

A couple of dozen ubiquitous think-tank acronyms are explained in a glossary that closes the book. Most of this readable volume is devoted to narratives of what in fact goes on at 10 or 15 of these lettered places. At the Institute for Land Combat they plan and game out wars of the future, doing with computer just what everyone believed the Prussian general staff was doing long ago with pencil and paper. In an Alexandria office building not far from the Pentagon they have thought up some 145 wars that might plausibly involve the U.S. One example is a desert war in North Africa with the U.S., Libya, Tunisia, Italy and West Germany fighting Algeria, the United Arab Republic and the U.S.S.R.

The Hudson Institute was founded by RAND heretics who left the sunny purlieus of Santa Monica in the name of the engaging legendary prophet Herman Kahn, half-droll and half-dread, to whom they still impute their most colorful irresponsibilities. The place is described factually, if somewhat implausibly, in one chapter. One of its recent specialties is the "flying think tank." Its teams flew and thought over Angola for 10 days in four light planes. Two volumes of papers came out of this incontrovertible overview (which was too high to see any fighting). Some specific ideas include damming the Congo, creating a wine industry, organizing safaris with "black hunters" and setting up a Mountie-like all-black elite border patrol that would always get its man. This report to a Portuguese manufacturer, complete with detailed scenarios, cost $100,000. It included a paper ascribed to Kahn himself. The prophet was not, of course, physically with the team on high. He dwells in still thinner air. Hudson has dropped most of the old operations-research tradition, with its heavy computing and rich tables of real or as-if data. The members of the Hudson Institute "often boast that its computing power is, with a gesture to the head, 'all up here.'" Somehow this "politically Rightist" ideological fiction factory finds apt clients. ABM and Vietnamization have had Hudsonian inputs.

The genre is still wider. There is room for the corporate study around Mr. Nader and the dissenting policy thought around Drs. Raskin and Waskow, just as for the semidetached Brookings Institution. There is the R&D supermarket of Arthur D. Little, where once they made a real silk purse—you can see it today in the Smithsonian—out of 100 pounds of sows' ears. That low yield hardly disposes of the adage, but it somehow seems to be the most unifying comment one can make on this typically American phenomenon, a collective but constrained search for reason applied in our competitive and corporate society.

RESTLESS EARTH: A REPORT ON THE NEW GEOLOGY, by Nigel Calder. Viking Press ($10). In his third book in the same novel and successful style the experienced, thoughtful English science writer once more presents us with a visually interesting, wide-ranging, remarkably up-to-date review of a field. His topic is both exciting and simple: it is the rebirth of geology, which since the "geopoetry" of Harry H. Hess in the early 1960's has undergone an unexpected rejuvenation, a growth in power hardly rivaled since an old toad last turned into a young prince. Geologists used to know all about some valley slope or small mountain range but had almost forgotten the diameter of the earth. Now they speak with confidence of connections between Cyprus and Newfoundland, or between the rise of the Himalayas and the origins of the coal deposits of Antarctica, or of the relations between Candide's Lisbon earthquake and the volcanic explosion of island Thera that blighted Minoan Crete. Readers will be at no loss to recognize the new geology; it has been spelled out in this magazine as it grew.

"The boldest notion of the new geology is very simple. It is that every great mountain chain on the planet has been produced by plate movements and usually by the destruction of an ocean." The continents ride like fragile rafts on the six major plates that build the shell of the earth, 40 miles thick. These plates travel, sliding on a pliant layer, nearly rigidly until they collide. New hot plate stuff wells up at the mid-ocean ridges, and sinks amid seismic friction to the depths again where plate edges collide, one plate "being swallowed under the margin of the other." The weaker, less dense continental material cannot submerge, and it largely wrinkles into mountains rather than sliding down into the abyss. India wrinkled into the Himalayas about 50 million years ago, in a violent collision with Tibet; the peninsula has not yet quite stopped moving. India drifted across a vanished ocean, speeding at a steady rate of a couple of inches a year.

The story is well told, and the photographs, diagrams, maps and reports of conversations on the issues with the men and women who make the science add clarity and a rare immediacy. The book is itself only a small part of a much larger event. It derives from a long television show, 90 minutes of science presented in color, the camera looking sometimes at real people who work at these problems day by day and sometimes at actual events and significant places: new lava flowing in Hawaii, ocean rifts gone ashore in Iceland or East Africa, the long straight seam of the San Andreas fault down the coast of California, the oldest rocks in the world in southwestern Greenland. The volume reflects the direct yet synoptic mode of exposition that has now been used for three sciences. Perhaps this travel-and-ask style works best of all for the new geology; the worldwide tour of this B.B.C. production team (Calder is the writer, of course) can bring the viewer-reader close to earth's nature and straight into the heart of the matter. The television shows are remarkably popular: the program on the new geology attracted more viewers than any other show so far exhibited on the U.S. National Educational Television network save one, a Buster Keaton old-film special. (The poll-makers took a quick recount, because the audience was double what they had expected for a documentary.) In addition to the American and British networks, Swedish, German and Australian television also helped to finance and have presented the program. Millions of people have seen and heard the images and narrative that lie behind and in Calder's book, an honest popularization of science with an unprecedented appeal.

One episode, both in the book and on the program, is worth retelling. It has been plain for some years now that water forced down a deep well into stressed rock can lubricate a fault, so that it releases its strain by small earthquakes. Pumping out water seems to reverse the effect, locking the slipping fault. Now, the Pacific plate "sidles northwards" through the San Francisco suburbs at a mean rate of two and a half inches a year. "No conceivable human force is going to stop that motion, which originates deep in the Earth" and is shared by the entire ocean floor from California to Japan. The rock drifts irregularly; just now it is "stuck in an ominous way." The slippage may be 13 feet in arrears of the general motion since the 1906 earthquake. Maybe the stress can be eased by drilling into the fault wells two and a half miles deep, three holes at a time spaced 500 yards apart. First you pump dry the two outer holes, thus locking the fault; then water pumped into the central hole should start local slippage, relieving the stress there but preventing any important motion beyond the "locks" of the two outer dry wells. Three by three, at a spacing of six miles, 500 boreholes at $1 million each might effectively stitch up the entire fault in time to spare us a predictable lethal disaster like Lisbon's. The therapy could be repeated every generation or so with the same wells. Such is the perspective of a new geology, with insight at last into the causes and nature of the deeper features of the earth.

INDEX OF ADVERTISERS

JULY 1972

AMERICAN OIL COMPANY 10, 11
Agency: D'Arcy-MacManus-Intermarco, Inc.

AMERICAN TELEPHONE AND TELEGRAPH COMPANY, THE 12
Agency: N. W. Ayer & Son, Inc.

BURROUGHS CORPORATION 2
Agency: Campbell-Ewald Company

CELESTRON PACIFIC, INC. 110
Agency: Baytron, Inc.

CHRYSLER CORPORATION, CHRYSLER-PLYMOUTH DIVISION ..... Inside Front Cover
Agency: Young & Rubicam, Inc.

DOVER PUBLICATIONS, INC. 119
Agency: James R. Flanagan Advertising Agency

EASTMAN KODAK COMPANY 47
Agency: Rumrill-Hoyt, Inc.

FORD MOTOR COMPANY 9
Agency: Grey Advertising, Inc.

FREEMAN, W. H., AND COMPANY 123

GCA CORPORATION 5
Agency: Culver Advertising, Inc.

GENERAL MOTORS CORPORATION, CHEVROLET MOTOR DIVISION 83
Agency: Campbell-Ewald Company

HEWLETT-PACKARD 114, 115
Agency: Richardson Seigle Rolfs & McCoy Inc.

JOHNS HOPKINS PRESS, THE 105
Agency: Welch, Mirabile & Co., Inc.

LIBRARY OF SCIENCE, THE 116
Agency: Henderson & Roll, Inc.

LILLY, ELI, AND CO. ..... Back Cover
Agency: Geer, DuBois & Co., Inc.

LUFTHANSA GERMAN AIRLINES 61
Agency: H. K. McCann Company mbH

MANPOWER INC. 101
Agency: Nader-Lief, Inc.

MARLEY COMPANY, THE 7
Agency: Valentine-Radford, Inc.

MERCEDES-BENZ OF NORTH AMERICA, INC. 61
Agency: Ogilvy & Mather Inc.

MILLER, HERMAN, INC. ..... 124, Inside Back Cover
Agency: Odiorne Industrial Advertising, Inc.

NATIONAL CAMERA, INC. 82
Agency: Langley Advertising Agency

NATIONAL CAR RENTAL SYSTEM, INC. 64
Agency: Richard K. Manoff Inc.

NEWPORT NEWS SHIPBUILDING & DRY DOCK COMPANY 1
Agency: Lawler Ballard Little Advertising

NIPPON KOGAKU K. K. 64
Agency: K & L Advertising Agency

OLYMPUS OPTICAL CO., LTD. 113
Agency: Fuji Agency, Ltd.

PAPERBACK LIBRARY 112
Agency: Herbert Arthur Morris Advertising, Inc.

POLAROID CORPORATION, THE 62, 63
Agency: Doyle Dane Bernbach Inc.

QUESTAR CORPORATION 49

RENAULT, INC. 48
Agency: Gilbert Advertising Agency, Inc.

SCRIBNER'S, CHARLES, SONS 118
Agency: March Advertising Inc.

UNITED AUDIO PRODUCTS, INC. 4
Agency: Ries Cappiello Colwell, Inc.

UNIVERSAL OIL PRODUCTS COMPANY 58
Agency: Campbell-Mithun, Inc.

WFF'N PROOF 82
Agency: Ad-Com Agency

BIBLIOGRAPHY

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

ANTISUBMARINE WARFARE AND NATIONAL SECURITY

FUNDAMENTALS OF SONAR. J. W. Horton. U.S. Naval Institute, 1959.

PRINCIPLES OF UNDERWATER SOUND FOR ENGINEERS. R. J. Urick. McGraw-Hill Book Company, 1967.

SIPRI YEARBOOK OF WORLD ARMAMENTS AND DISARMAMENT 1969/70. Stockholm International Peace Research Institute. Humanities Press, 1971.

ASW—THE CRUCIAL NAVAL CHALLENGE. Captain Robert H. Smith, USN ret., in *U.S. Naval Institute Proceedings*, pages 126–141; May, 1972.

MISSILE SUBMARINES AND NATIONAL SECURITY. Herbert Scoville, Jr., in *Scientific American*, Vol. 226, No. 6, pages 15–27; June, 1972.

THE X-RAY SKY

POLARIMETER FOR CELESTIAL X RAYS. H. W. Schnopper and K. Kalata in *The Astronomical Journal*, Vol. 74, No. 7, pages 854–858; September, 1969.

PRECISE LOCATION OF SAGITTARIUS X-RAY SOURCES WITH A ROCKET-BORNE ROTATING MODULATION COLLIMATOR. H. W. Schnopper, H. V. Bradt, S. Rappaport, E. Boughan, B. Burnett, R. Doxsey, W. Mayer and S. Watt in *The Astrophysical Journal*, Vol. 161, No. 3, Part 2, pages L161–L167; September, 1970.

SOFT X RAYS FROM THE VICINITY OF THE NORTH POLAR SPUR. A. Bunner, P. Coleman, W. Kraushaar and D. McCammon in *The Astrophysical Journal*, Vol. 172, No. 2, Part 2, pages L67–L72; March, 1972.

LEWIS CARROLL'S LOST BOOK ON LOGIC

SYMBOLIC LOGIC [Part I] AND THE GAME OF LOGIC. Lewis Carroll. Dover Publications, Inc., 1958.

A HISTORY OF FORMAL LOGIC. I. M. Bocheński, translated and edited by Ivo Thomas. University of Notre Dame Press, 1961.

THE PARADOX OF THE LIAR. Edited by Robert L. Martin. Yale University Press, 1970.

READINGS ON LOGIC. Edited by Irving M. Copi and James A. Gould. The Macmillan Company, 1972.

THE CHEMICAL ELEMENTS OF LIFE

TRACE ELEMENTS IN BIOCHEMISTRY. H. J. M. Bowen. Academic Press, 1966.

CONTROL OF ENVIRONMENTAL CONDITIONS IN TRACE ELEMENT RESEARCH: AN EXPERIMENTAL APPROACH TO UNRECOGNIZED TRACE ELEMENT REQUIREMENTS. Klaus Schwarz in *Trace Element Metabolism in Animals*, edited by C. F. Mills, E. & S. Livingstone, 1970.

THE PROTEINS: METALLOPROTEINS, VOL. V. Edited by Bert L. Vallee and Warren E. C. Wacker. Academic Press, 1970.

CERULOPLASMIN: A LINK BETWEEN COPPER AND IRON METABOLISM. Earl Frieden in *Bioinorganic Chemistry*, Advances in Chemistry Series 100. American Chemical Society, 1971.

TRACE ELEMENTS IN HUMAN AND ANIMAL NUTRITION. E. J. Underwood. Academic Press, 1971.

THE TOKAMAK APPROACH IN FUSION RESEARCH

PHYSICS OF HIGH TEMPERATURE PLASMAS. George Schmidt. Academic Press, 1966.

ENGINEERING FEASIBILITY OF CONTROLLED FUSION: A REVIEW. D. J. Rose in *Nuclear Fusion*, Vol. 9, No. 3, pages 183–203; October, 1969.

TURBULENCE IN TOROIDAL SYSTEMS. B. B. Kadomtsev and O. P. Pogutse in *Reviews of Plasma Physics: Vol. 5*, edited by M. A. Leontovich. Consultants Bureau, 1970.

DEPRIVATION DWARFISM

GROWTH FAILURE IN MATERNAL DEPRIVATION. Robert Gray Patton and Lytt I. Gardner. Charles C Thomas, Publisher, 1963.

EMOTIONAL DEPRIVATION AND GROWTH RETARDATION SIMULATING IDIOPATHIC HYPOPITUITARISM, II: ENDOCRINOLOGIC EVALUATION OF THE SYNDROME. G. F. Powell, J. A. Brasel, S. Raiti and R. M. Blizzard in *The New England Journal of Medicine*, Vol. 276, No. 23, pages 1279–1283; June 8, 1967.

SHORT STATURE ASSOCIATED WITH MATERNAL DEPRIVATION SYNDROME: DISORDERED FAMILY ENVIRONMENT AS CAUSE OF SO-CALLED IDIOPATHIC HYPOPITUITARISM. Robert Gray Patton and Lytt I. Gardner in *Endocrine and Genetic Diseases of Childhood*, edited by L. I. Gardner. W. B. Saunders Company, 1969.

GROWTH HORMONE IN NEWBORN INFANTS DURING SLEEP-WAKE PERIODS. Bennett A. Shaywitz, Jordan Finkelstein, Leon Hellman and Elliot D. Weitzman in *Pediatrics*, Vol. 48, No. 1, pages 103–109; July, 1971.

EXPERIMENTS IN READING

THE PSYCHOLOGY AND PEDAGOGY OF READING. Edmund Burke Huey. The M.I.T. Press, 1968.

CLUES TO A LETTER'S RECOGNITION: IMPLICATIONS FOR THE DESIGN OF CHARACTERS. Paul A. Kolers in *The Journal of Typographic Research* (now *Visible Language*), Vol. 3, No. 2, pages 145–168; April, 1969.

BASIC STUDIES ON READING. Edited by Harry Levin and Joanna P. Williams. Basic Books, Inc., 1970.

ESCAPE RESPONSES IN MARINE INVERTEBRATES

PREDATOR RECOGNITION AND ESCAPE RESPONSES OF SOME INTERTIDAL GASTROPODS IN THE PRESENCE OF STARFISH. Theodore Holmes Bullock in *Behaviour*, Vol. 5, Part 2, pages 130–140; 1953.

GASTROPOD DEFENSIVE RESPONSES AND THEIR EFFECTIVENESS IN REDUCING PREDATION BY STARFISHES. Howard M. Feder in *Ecology*, Vol. 44, No. 3, pages 505–512; Summer, 1963.

ASPECTS OF ASTEROID BIOLOGY. Howard M. Feder and Aage Møller Christensen in *Physiology of Echinodermata*, edited by Richard A. Boolootian. Wiley-Interscience, 1966.

FEEDING BEHAVIOR OF ASTEROIDS AND ESCAPE RESPONSES OF THEIR PREY IN THE PUGET SOUND REGION. Karl P. Mauzey, Charles Birkeland and Paul K. Dayton in *Ecology*, Vol. 49, No. 4, pages 603–619; Early Summer, 1968.

MATHEMATICAL GAMES

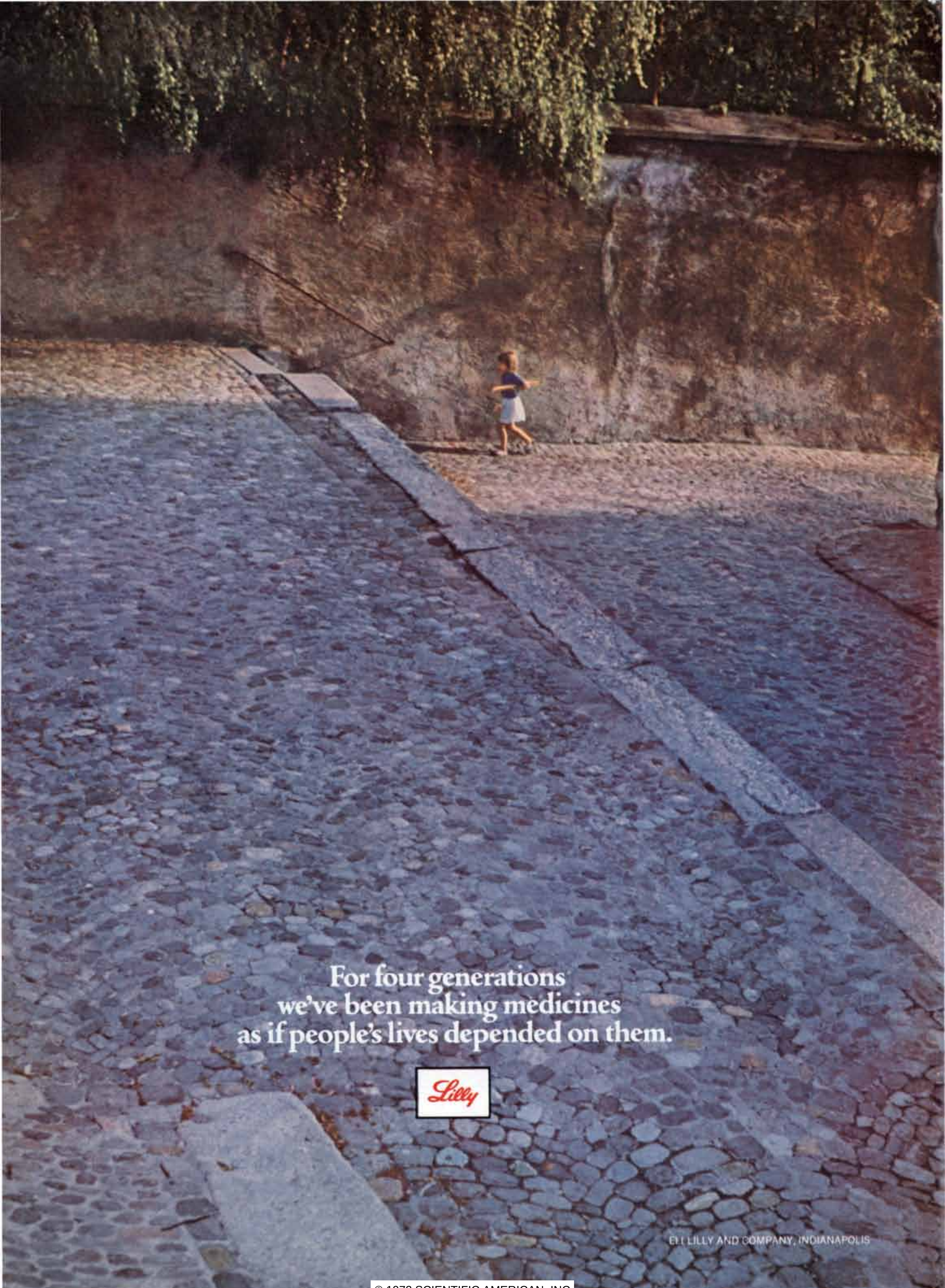
SCARNE ON CARD TRICKS. John Scarne. Crown Publishers, Inc., 1950.

MATHEMATICS, MAGIC AND MYSTERY. Martin Gardner. Dover Publications, 1956.

MATHEMATICAL MAGIC. William Simon. Charles Scribner's Sons, 1964.

THE AMATEUR SCIENTIST

WIND-TUNNEL TESTING. Alan Pope. John Wiley & Sons, Inc., 1947.


For four generations
we've been making medicines
as if people's lives depended on them.
Lilly
ELI LILLY AND COMPANY, INDIANAPOLIS