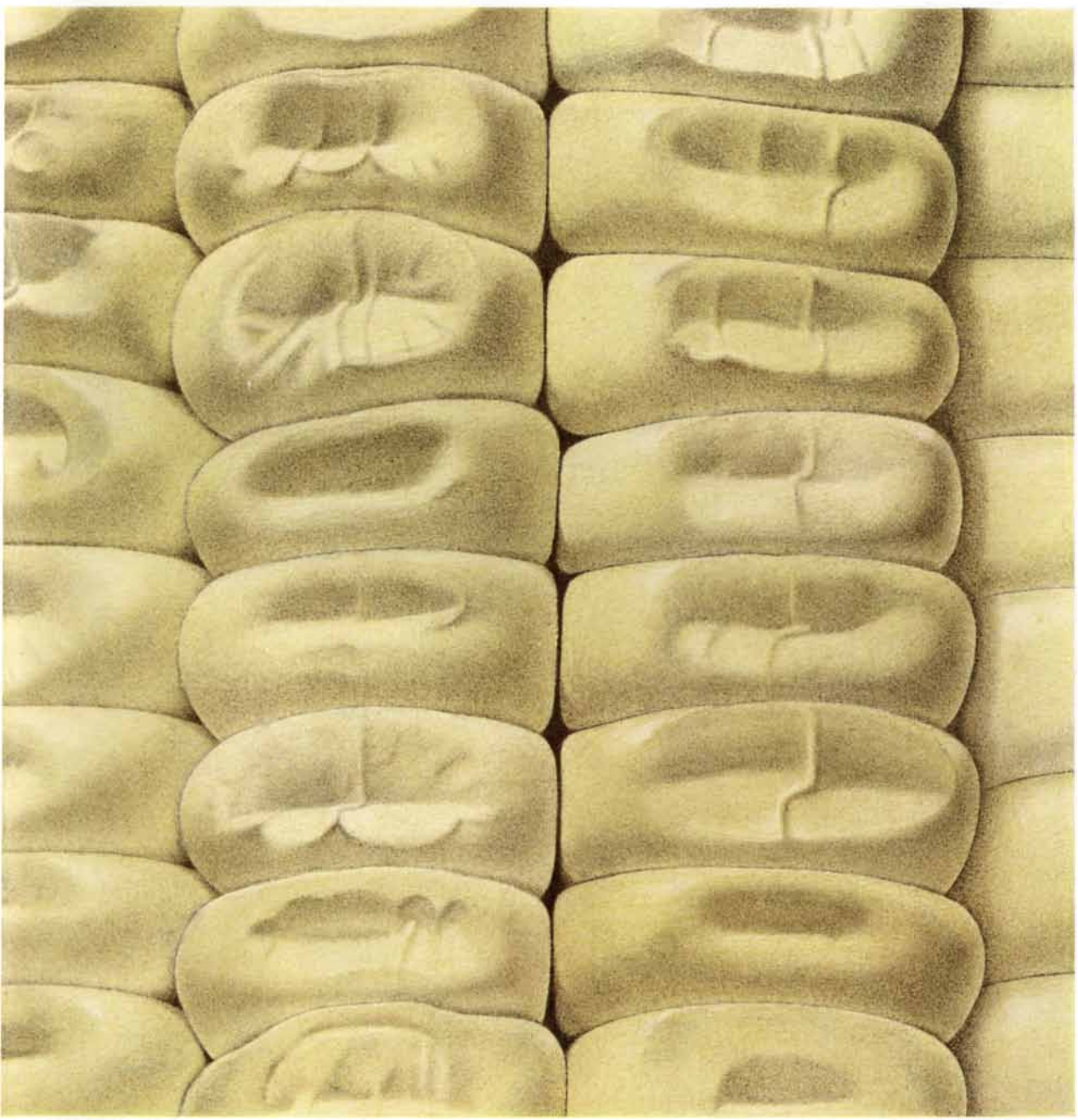
SCIENTIFIC AMERICAN



HIGH-LYSINE CORN

ONE DOLLAR

August 1971

You have problems. We have solutions.

Auto pollution is the gasoline problem UOP is cracking now, and it may have some projectable solutions for you.

We've developed two ways of combating the growing automotive pollution menace. Both reflect the innovativeness that made us pioneers in petroleum refining.

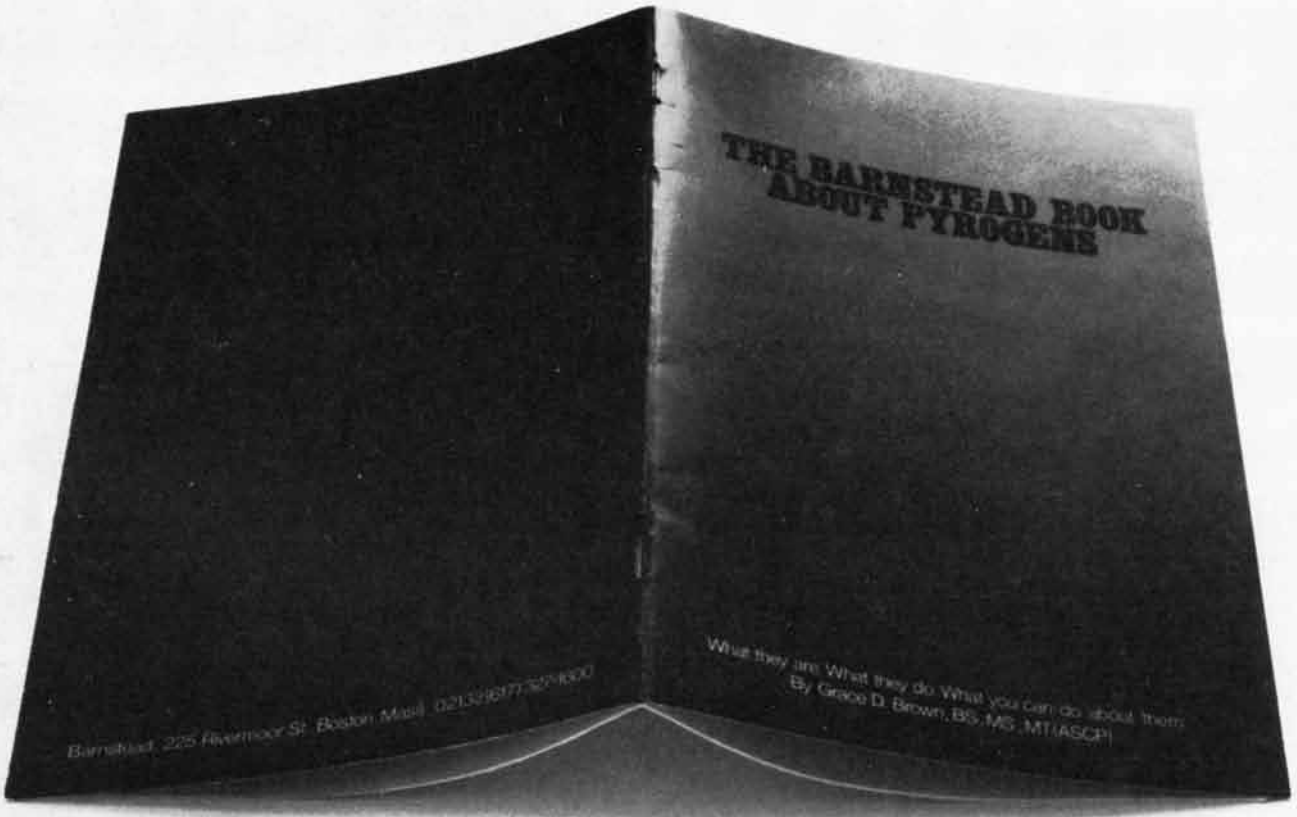
First, our catalytic converter, a muffler-like device for your car that can reduce automotive pollutants up to 90% when used with lead-free gas.

Second, we have refining processes to speed the economical production of high octane unleaded gasoline. Plus the capabilities to design and construct plants anywhere in the world.

Our scientists and researchers are helping refiners process a cleaner product. They made these solutions work for you and they can help with other problems of environmental control. But, first, they need to know what problems you want solved.

UOP serves the petroleum and automotive industries through these divisions: (a) *UOP Research Center*, development of catalytic converters for all cars. (b) *Processes by UOP*, advanced refining methods to help refiners solve the lead problem. (c) *Engineered construction* by Procon, Inc. to solve construction problems throughout the petroleum and petrochemicals industry. **For solutions to pollution problems, write, call UOP (Universal Oil Products Company), UOP Plaza, Des Plaines, Ill. 60016.**

We put this little book together because pyrogens can be a lot more than a pain in the neck.



We paid Grace D. Brown of totally independent G.B. Laboratories, Inc., Brookline, Mass., to do it.

To put down everything anybody knows about pyrogens.

What they are.

The damage they can do.

And what you can do about them.

Despite the fact that this little book proves that the best way to make sure you're not going to have a pyrogen problem is to run your water through one of our new TD-15 stills, it's much more than a Barnstead commercial message.

It's the last word on pyrogens.

And if you're making eye drops, mouthwash, toothpaste, pancake, lipstick, eye liner, intravenous or other medical solutions, anything that goes on or into a human being it's the end of your search to make it safer.

We're not in the printing business so we've got to charge you a 25¢ handling charge for handing your book to you.

But any information you want about our TD-15's is free.

Some of the best things you can get in life still are.

Even if little books about what you can do with it aren't.

Mr. Norman McLellan

BARNSTEAD
SYBRON CORPORATION

225 Rivermoor St./Boston, Mass. 02132/(617) 327-1600

☐ Enclosed is my 25¢ for The Barnstead Book About Pyrogens.
☐ Please send me my free information on your TD-15 stills.

Name_____

Title_____

Company_____

Address_____

City_____ State_____ Zip_____

In the minds of many, modern technology has created a monster.

The computer.

We've all heard the stories about people making, say, a $30 purchase. And then being billed for $3,000 by the computer.

Nonsense.

The danger is not that the computer makes mistakes, but that human errors remain uncorrected while the machine rolls on, compounding them.

Computers are literal minded. They must be correctly instructed to help us in the solution of problems. They do exactly what they are told. Not what they ought to have been told.

The computer is man's assistant. Not his replacement.

The unaided human mind needs help to cope successfully with the complexity of our society.

Intellectual aids, such as computers, will not only increase the skill of our minds, but leave more time for human creativity by freeing man of burdensome routine tasks.

Do we really believe that our achievements in space could have been accomplished without computer assistance?

Do we really believe that we can function efficiently in our complex modern environment without computer assistance?

The answer, of course, is obvious.

In truth, the invention of the computer can be compared with the invention of the printing press.

Engineers engaged in the development of computer systems are convinced that over the next decade it is possible to develop networks of interconnected computer systems capable of offering a wide variety of services to the public.

By necessity, one-way mass communications — radio, television—deal with a common denominator of entertainment. This situation can be changed by developing computer-based systems that offer each individual an almost unlimited range of entertainment and information. Each individual will select what he wants, and to how great a depth he wants to delve into the areas in which he is interested.

At his choice of time.

Apply this principle to education.

What it amounts to is individualized instruction. To meet simultaneously the needs of many students.

From a practical standpoint, limits to excellence in education are almost purely economic.

The computer provides a solution by performing high quality instruction for large numbers of students, economically.

Our goal is to make it possible for a teacher to provide individual guidance to many students, instead of few.

Yet, computer-assisted instruction is not a concept which has been enthusiastically embraced by all. There are many who feel that the computer will replace teachers.

Not so.

This interpretation implies mechanizing, rather than personalizing, education.

Everywhere in our lives is the effect and promise of the computer.

Its ability to predict demand makes it possible to apply the economies of mass production to a wide variety of customized products.

It will allow for the use of a computer terminal device for greater efficiency in home shopping and much wider diversity in home entertainment.

It can be a safeguard against the boom and bust cycle of our economy.

In short, the computer means accuracy, efficiency, progress.

The computer affords us the way to store knowledge in a directly usable form—in a way that permits people to apply it without having to master it in detail.

And without the concomitant human delays.

The computer is indicative of our present-day technology —a technology which has advanced to such an extent that man now is capable, literally, of changing his world.

We must insure that this technological potential is applied for the benefit of all mankind.

If you are interested in a comprehensive index of over 1100 technical papers published recently by RCA scientists and engineers, let us know.

Write to: Mr. D. L. Crosier, RCA, Bldg. 2-2, Camden, New Jersey 08102.

Of course, we're an equal opportunity employer.

RCA

ARE THEY FOR US OR AGAINST US?

Established 1845 **SCIENTIFIC AMERICAN** August 1971 Volume 225 Number 2

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.

THE COVER

The painting on the cover shows the kernels on an ear of high-lysine corn: a strain that largely corrects for a deficiency of the amino acid lysine in normal corn (see "High-Lysine Corn," page 34). Lysine is an essential amino acid in the nutrition of man and other nonruminant animals such as the pig and the chicken; normal corn alone cannot meet the nutritional requirements of such animals. High-lysine corn can nearly do so by itself. It is produced by breeding into normal corn strains recessive genes that favor the synthesis of the protein glutelin, which is rich in lysine, over the protein zein, which is not. The endosperm in the kernels of the present prototypes of high-lysine corn has a characteristic floury appearance, which is reflected in the light yellow coloration of the kernels in the painting.

THE ILLUSTRATIONS

Cover illustration by Lorelle A. Raboni

Page	Source
16	Jim Egleson
17	Homer Page, World Health Organization (*top*); U.S. Public Health Service (*bottom*)
18–21	Jim Egleson
22	George F. Bass, University Museum, University of Pennsylvania
24–25	Jerome Kuhl
26–32	Eric Mose
33	Jerome Kuhl
34	Ben Rose
35–36	Tom Prentiss
37	Uheng Khoo and F. L. Baker, Northern Regional Research Laboratory, U.S. Department of Agriculture
38–41	Tom Prentiss
42	Ben Rose
48	Stephen Karakashian, State University of New York at Old Westbury
50–52	Enid Kotschnig
53	A. V. Grimstone, University of Cambridge (*top*); William T. Hall, National Institutes of Health (*bottom*)
54	Richard D. Allen, University of Hawaii (*top*); Enid Kotschnig (*bottom*)
56	Enid Kotschnig
57	Sigmund Socransky, Forsyth Institute (*top*); Toichiro Kuwabara, Harvard Medical School (*bottom*)
62	National Aeronautics and Space Administration
63	Dan Todd
64	National Aeronautics and Space Administration
65–73	Dan Todd
75	Paul Weller
76–81	Bunji Tagawa
82–89	Allen Beechel
93–105	Alan D. Iselin
106–115	Roger Hayward

SCIENCE/SCOPE

The health of America's crops, forests and rivers will be checked every 18 days next year by a new scanning device aboard NASA's Earth Resources Technology Satellite. The 100-lb. optical-mechanical instrument, called a Multi-spectral Scanner, was developed by Hughes. It is designed to detect and record the different "signatures" of the solar energy emitted by all objects on Earth and to convert them into photo-like images that will show the condition of various natural resources.

The synchronous communications satellite's first decade was featured in the Hughes display at the Telecom '71 exhibition staged by the International Telecommunications Union in Geneva, Switzerland, June 17-28. Included were: the first synchronous satellite, shown at the 1961 Paris Air Show by the Hughes team credited with the original concept; a full-scale model of Anik I, Canada's new domestic satellite; and third-scale models of all others built by Hughes -- from Syncom, world's first, which was launched in 1963, to the giant Intelsat IV, which began commercial service March 26.

The U.S. Army Safeguard System Command, Huntsville, Ala., recently awarded letter contracts to three companies for contract definition leading to the competitive selection of a prime contractor for a hardsite defense prototype demonstration program. Hughes has teamed with Boeing and System Development Corp. The work will be performed over a five-month period.

Airborne radar transmitter design engineers are needed now at Hughes. Must have specific fire-control-system, doppler, pulse-compression, microwave, and power-supply experience. Also: solid state microwave engineers with experience ranging from UHF to millimeter frequencies, and in the design and use of related circuits. Both positions require accredited degree, 3 years of specific experience, and U.S. citizenship. Write: Mr. Robert A. Martin, Hughes Aerospace Engineering Divisions, 11940 W. Jefferson Blvd., Culver City,CA 90230. An equal opportunity M/F employer.

A new traveling wave tube for Canada's Anik I domestic synchronous communications satellite will operate even more reliably and with higher efficiency than previous Hughes TWTs. It is expected to operate for more than 12 years, compared with the six-month life expectancy of the Syncom II TWT (which, however, is still operable after eight years). Hughes TWTs have also been used on all the Intelsat, ATS, TACSAT, Mariner, and Lunar Orbiter satellites and the Surveyor and Apollo spacecraft. Their record to date: 100 years in space without a relevant failure.

The first tri-service validation of an Air Force contractor's program performance measurement system was won by Hughes recently on the cost-schedule control system for the Maverick missile program. Maverick -- a TV-guided air-to-ground missile -- is being developed under a "total package procurement" contract. It has completed flight tests by Hughes and is now in USAF's Category II flight test.

LETTERS

Sirs:

In their article on quasars [SCIENTIFIC AMERICAN, May] Maarten Schmidt and Francis Bello extrapolate "with reasonable confidence" from data on 20 tentatively identified quasars to a total of about 15 million unobserved objects. However, they regard the evidently too bold hypothesis of ejection of quasars from a gravitational collapse at the center of our galaxy as quite unlikely. This is, of course, a matter of personal opinion, but it is hard to understand why the same easily refuted objections to local origin of quasars are cited once again.

The objection that blue-shifted quasars have not been detected from other galaxies is not valid if local quasars are quite close. Proper motion data for the quasars of lowest red shift, such as 3C 273, are consistent with a distance of only a few hundred thousand light-years (J. Terrell, *Physical Review Letters,* Vol. 21, page 637; 1968). In this case they would have perhaps 10,000 to 100,000 times the brightness of the sun and would need only a few solar masses. That these parameters are similar to those of Type O stars has been pointed out by Siu-Kay Luke (to be published).

Such objects would be faint enough if ejected even by nearby galaxies to make a red-shift (or blue-shift) measurement exceedingly difficult. Only the radio emission from large numbers of such objects would be readily detectable—and may be the explanation of the puzzling radio sources near many radio galaxies.

The objection to this hypothesis on energetic grounds is equally invalid for such local quasars. The ejection of even a million such objects of a few solar masses at relativistic speeds would put no particular strain on the resources of a galactic center. On the contrary, it is the cosmological hypothesis that runs into energy and mass problems.

The recent observation by James E. Gunn (*The Astrophysical Journal,* Vol. 164, *Letters to the Editor,* page 113; 1971) of a quasar, PKS 2251 + 11, "superposed on a small, compact cluster of galaxies" meets a conspicuous need of the cosmological quasar hypothesis. The data from which the red shift of one of these galaxies is found to be nearly the same as that of the quasar ($z = 0.323$) would not be considered acceptable proof, however, if the red shift were not already known. If the red shift is correctly estimated, then the cluster of galaxies is unusually dim and amazingly compact, which is no doubt why H. C. Arp (*The Astrophysical Journal,* Vol. 162, page 811; 1970) concluded that this same quasar was *not* part of a cluster of galaxies.

A very recent discovery that may lend support to local quasars is the report by C. A. Knight *et al.* (*Science,* Vol. 172, page 52; 1971) that two radio sources associated with the quasar 3C 279 are apparently separating at a speed in excess of nine times the speed of light! This is, of course, a measurement of change of angle, and the assumption of a smaller distance than that given by Hubble's law would greatly ease the situation.

Local origin of observed quasars seems the most physically reasonable explanation of much of the observational data, as I have pointed out elsewhere, for instance in *Science* (Vol. 154, page 1281; 1966) and in *The Astrophysical Journal* (Vol. 147, page 827; 1967). If it were not for uncooperative observations, the evidently more glamorous cosmological quasars would have been fully accepted long since.

JAMES TERRELL

Los Alamos Scientific Laboratory
Los Alamos, N.M.

Sirs:

A survey of a small part of the sky shows that to magnitude 19½ there are about five quasars per square degree (A. Braccesi *et al.*, *The Astrophysical Journal,* Vol. 152, *Letters to the Editor,* page 105; 1968). Considering the steep increase in the number of quasars with magnitude, a total number of one million over the sky is a conservative estimate. If the average mass of the individual quasar is 1,000 solar masses (J. Terrell, *Science,* Vol. 154, page 1281, 1966; *The Astrophysical Journal,* Vol. 147, page 827, 1967; *Science,* Vol. 156, page 265, 1967), the total kinetic energy of the explosion equals 10^{63} ergs. This corresponds to the *total rest-mass energy* of 10^9 solar masses, that is, all the mass within 300 light-years from the nucleus of our galaxy. I mentioned in our May article an individual mass of 10,000 solar masses, which is a minimum estimate derived from the observed properties of the emission-line spectra of quasars (M. Schmidt, Volume IX of *Stars and Stellar Systems,* edited by A. and M. Sandage, to be published).

These excessive energy requirements are based on an age of the explosion of five million years, which is the minimum age corresponding to a maximum proper motion for 3C 273 of .002 second of arc per year. The distance of 3C 273 would be 600,000 light-years and quasars with larger red shifts would be at distances of millions of light-years. Hence some of "our" quasars would be close to neighboring galaxies, and we might expect to see some of "their" quasars exhibiting blue shifts, if they existed.

Dr. Terrell has attempted to alleviate the above objections by revising the age of the explosion to 1.55 million years (*Physical Review Letters,* Vol. 21, page 637; 1968). Since in his discussion the energy required depends on the cube of the distance of the quasars, the total energy requirement is reduced by a factor of 30. However, the reduced age is derived by using the least accurate of two proper-motion determinations of 3C 273 and allowing an error of two standard deviations in a preferred direction. The adopted proper motion deviates more than five standard deviations from the more accurate determination. This approach is artificial, and it is hard to understand why Dr. Terrell considers objections to the local origin of quasars "easily refuted." Parenthetically, the minimum quasar mass of 10,000 solar masses derived from the emission-line spectrum is independent of the age of the explosion; hence on that basis the total energy requirement cannot be reduced by lowering the age.

The apparent superrelativistic increase of the angular separation of two radio components of the quasar 3C 279 can be understood on the basis of ejection at relativistic speed if the direction of ejection makes a not too large angle with our line of sight. A similar situation was not considered objectionable in the local theory of quasars a few years ago (J. Terrell, *Physical Review Letters,* Vol. 21, page 637; 1968). Whether or not the phenomena seen in 3C 279 constitute an argument against the cosmological hypothesis depends on further observations of this source and other sources, the results of which are awaited with interest.

MAARTEN SCHMIDT

Hale Observatories
Pasadena, Calif.

Sirs:

One may forgive Riley D. Woodson ["Cooling Towers," SCIENTIFIC AMERI-

CAN, May] for the strong bias to the cooling problems of power stations, since this is the major field of application of cooling towers. However, as a consequence of this bias the article may create the wrong impression that cooling with air is *per se* more expensive than cooling with water. This is so *only* when the process outlet temperature is close to the ambient temperature (power-station condenser), in which case the disadvantage that "dry" systems cannot cool below ambient temperatures becomes rather serious.

For process outlet temperatures more than about 5 degrees Celsius above the ambient temperature, cooling with air can usually be done at lower *total* cost than cooling with water. In the process industry the alternative for an evaporative cooling tower is not a dry cooling tower in a cooling water system but direct cooling with air using air-cooled heat exchangers *in the process lines.* Cost comparisons must be made on a complete system and not on a single item (the cooling tower).

W. G. B. MANDERSLOOT

National Chemical Research Laboratory
Pretoria, South Africa

Scientific American, August, 1971; Vol. 225, No. 2. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and 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 Allan Wittman, Advertising Manager, 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

AUGUST, 1921: "The end of a modern war is certain to leave the victors only less impoverished than the vanquished. Moreover, it is the irony of fate that in respect to the late war the nations that won the decision should today find themselves borne down by military burdens that just about equal the cost of the reparations they have imposed. If we, in common with Great Britain, France and Japan, continue to sink enormous sums into the maintenance of fleets and armies on a scale that is out of all proportion to the legitimate international police necessities of these nations, we are placing ourselves under a financial burden that is practically equal to that which has been assumed by Germany. Fortunately there is to be held in this country a congress whose primary object will be the amicable settlement of outstanding international questions with a view to reasonable disarmament. The success of that momentous gathering, whose first session will take place in Washington on Armistice Day, will depend more than anything on the degree to which dissimulation gives place to a sincere desire to look impartially at both sides of every question, and on a resolute determination by all concerned to leave suspicion and fear outside the council chamber."

"In the ruthless destruction of our forests and the extravagant and wasteful methods by which we are using up the natural resources of America we have been following a policy that has been truly described as one in which it is a case of 'every man for himself and the Devil take the hindmost.' The most discouraging fact about the whole situation is that, in spite of endless warnings and the carefully prepared governmental statistics showing the rapid depletion of our resources, particularly of our forests, nobody seems to be very much disturbed and the movement to correct this abuse is apparently making very small headway."

"A French engineer, H. M. Melot, has put out an invention that from its form he calls the propelling trumpet. Capable of application to all sorts of vehicles, it is designed primarily for the airplane. The apparatus consists of a number of tubes ending with trumpet-like flares or nozzles. These are arranged in series in connection with a combustion chamber, where an explosive mixture of air and fuel is ignited as in ordinary engine practice. The exhaust gases of the combustion chamber are discharged into the series of nozzles. Both the pipe that effects this operation and the nozzles themselves are carefully designed to cause the expansion of the gases to occur under the best circumstances. It is the velocity of the exhaust and the velocity of expansion that, through reaction against the external air, drives the machine forward. The exhaust gases are discharged at a velocity of from 1,200 to 1,500 yards per second; at the entrance to each nozzle a certain amount of the outside air is drawn in and surrounds the jet of exhaust gas as perfectly as possible. The gas therefore gives up a part of its velocity to the air and causes a powerful suction action at the entrance to each nozzle."

"We have no doubt that on numerous occasions disgruntled ball players have failed to exert their efforts to win or have actually done what they could to betray their club. But concerted throwing of games for money, if we are to credit the testimony offered in the case of the Chicago baseball scandal, has been put on a business basis never before attained. A new crime has been invented, and the legislatures of numerous states have responded by defining the offense and the punishment."

SCIENTIFIC AMERICAN

AUGUST, 1871: "Alfred Russel Wallace's new book, *The Action of Natural Selection on Man: The Development of Human Races under the Law of Selection; the Limits of Natural Selection as Applied to Man,* is a significant addition to the already profuse literature on the subject of natural selection. Mr. Wallace agrees with Mr. Darwin that natural selection has played an important part in the development of the various races of men, but he denies that it alone accounts for the facts of man's present existence, his moral nature, etc. As a writer, Mr. Wallace is not inferior to Mr. Darwin; as a thinker, he ranks high, and his arguments are not of a kind that can be easily refuted."

"The story of the explosion of the boiler of the steam ferry boat *Westfield,* plying between Staten Island and New York, has been conveyed to every corner of the land. As yet the coroner's inquest is not completed, and the evidence scarcely warrants an opinion as to the cause of the dreadful disaster. It is, however, revealed that the boiler was old and patched, that steam pressure was carried above that authorized by the license, and that the engineer was away from his post at the time of the explosion and had been so at least five minutes, the pressure being 27 pounds when he left it—two pounds above the licensed maximum. How much the pressure rose during his absence is not known, but that it was enough to precipitate a large number of persons into eternity and grievously burn and maim still more is certain. There may have been negligence on the part of the inspector; that there was culpable neglect on the part of the company and employees is sure. A patched boiler is not necessarily an unsafe one, but patched principles are always unsafe. And it is because of looseness in the administration of law and the want of enforced regard to the public welfare on the part of railway and steamboat companies that the waters of New York Bay were filled with dead and wounded on the last Sunday in July, 1871."

"Professor Tyndall's *Fragments of Science for Unscientific People* are not all of them easy reading. You cannot run through them as you can through a lady writer's romance, although certain portions of them are more sensational than the most sensational novel. There is never an attempt to hide rough ground under the flowers of rhetoric or the sticks and straws of verbiage. Wherever there are hard places to be traversed, Professor Tyndall tells you they are and will be hard, and advises you to gird up your loins accordingly. When the trying bit of climbing is accomplished, 'thus patiently you have accompanied me over a piece of exceedingly difficult ground; and I think as a prudent guide we ought to halt upon the eminence we have now attained. We might go higher, but the boulders begin here to be very rough. At a future day, we shall, I doubt not, be able to overcome this difficulty, and to reach together a higher elevation.' Such frankness, combined with such lucidity, renders the reading of Professor Tyndall's works a mental tonic."

You can now reserve Gilroy Roberts' newest limited edition bird sculptures. In solid sterling silver. Deadline: Tuesday, August 31st.

STERLING SILVER GOSHAWK MEDAL SHOWN 2⅓ TIMES ACTUAL SIZE

The sterling silver Goshawk medal (shown opposite — 2½ times actual size) is one of five new bird sculptures, by the famed artist Gilroy Roberts, that will be issued in the form of solid sterling silver art medals.

All five of these medals are shown below in their actual size—2″ in diameter. Each is a masterpiece in its own right—a delicate work of art by Gilroy Roberts, whose portrait of John F. Kennedy graces the U.S. half dollar. (Mr. Roberts, a master engraver as well as a superb sculptor personally supervises translation of each sculpture into the medium of solid sterling silver.)

Each medal is minted with a beautiful proof finish by the Franklin Mint, the world's foremost private mint. Struck on heavy sterling silver planchets, weighing at least one thousand grains, these medals are coined in individually limited editions — with each edition permanently limited to *one* sterling silver specimen for Mr. Roberts' personal collection, *one* for the Franklin Mint archives, and *one* for each Patron whose order is placed before the strikings begin.

Thus, the total edition of each bird sculpture is completely "sold out" before the first piece is minted. *No additional pieces are struck for future sale.* After each Patron receives the sterling silver specimen struck expressly for him, the dies for that design are destroyed.

As a result, future collectors who want to acquire these designs must obtain them from Patrons who are willing to sell part or all of their collections, at prices determined by the laws of supply and demand.

You can now become a Patron of this program by ordering one or more of the sterling silver proof medals to be issued at this time.

Patrons of this newest group of five Roberts Birds must place their orders by AUGUST 31, 1971. No orders postmarked later than that date can be accepted. The original issue price is $20.00 for each bird or $90.00 for all five. Patrons may order one, two, three, four, or all five of these new Roberts Birds, but no more than one specimen of each design.

SKYLARK

GOSHAWK

EUROPEAN ROBIN

WOODCOCK

ALBATROSS

Each medal will be shipped to you encased in this handsome Lucite easel display.

ROBERTS BIRDS PATRON'S APPLICATION AND ORDER FORM

valid only if received in an envelope postmarked by AUGUST 31, 1971.

The Franklin Mint
Franklin Center, Pa. 19063

Please enroll me as a Patron of Roberts Birds and enter my order for the following proof-quality specimens, to be struck in solid sterling silver expressly for my personal collection:

☐ SKYLARK	$20.00
☐ GOSHAWK	$20.00
☐ EUROPEAN ROBIN	$20.00
☐ WOODCOCK	$20.00
☐ ALBATROSS	$20.00
☐ The Complete set of five	$90.00

Total of order $_____

Add your state sales tax _____

Remittance enclosed $_____

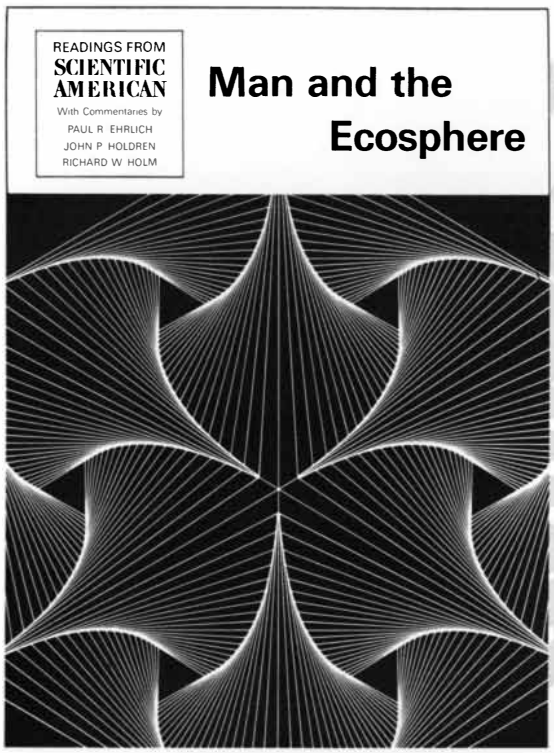
Mr.
Mrs.
Miss _____

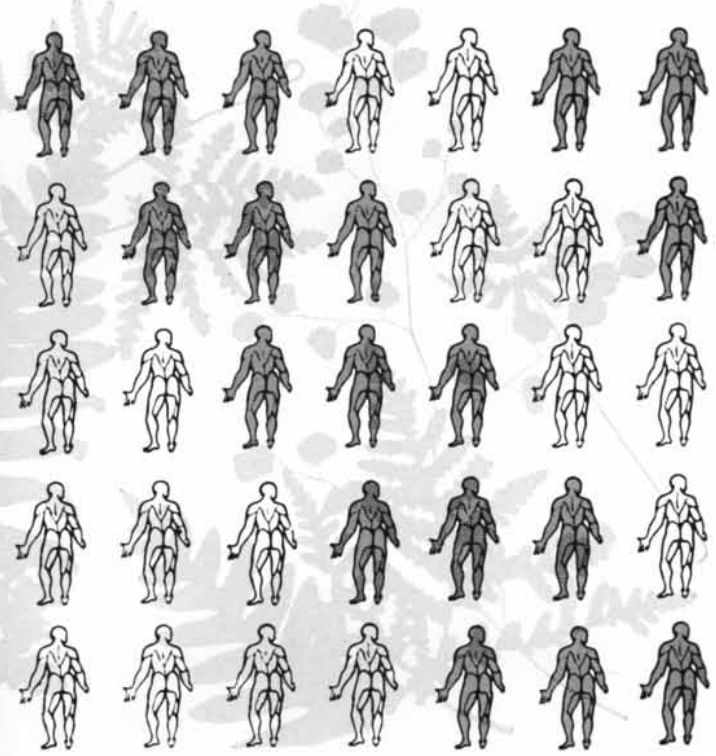
PLEASE PRINT CAREFULLY

Address _____

City _____ **State** _____ **Zip** _____

The Franklin Mint is an independent, publicly-owned corporation and is not affiliated with the U.S. Mint





Man and the Ecosphere

Readings from SCIENTIFIC AMERICAN

With Commentaries by
PAUL R. EHRLICH, Stanford University,
JOHN P. HOLDREN, Lawrence Radiation Laboratory,
and RICHARD W. HOLM, Stanford University

1971, 307 pages, 236 illustrations (129 in color),
Clothbound $11.00, paperbound $5.75

"We hope that these articles, together with the commentaries that introduce them, convey a sense of the dimensions of man's impact on the ecosphere and, correspondingly, of the fundamental sociological and technological changes required to alleviate that impact. . ." —From the Preface

A new solution to the problem of suitable supplementary reading for courses in environmental studies is provided in this book by two biologists and a physicist. They have organized 27 articles from SCIENTIFIC AMERICAN into an integrated presentation that illuminates the history of man's relationship with his environment. In each introduction to the four sections of the book, they offer extensive commentary on the articles and bring into focus our most pressing ecological problems.

Man and the Ecosphere will uniquely complement other books in ecology and will deserve a place on every concerned reader's bookshelf.

SECTIONS: The Ecosphere and Preindustrial Man. Limits Rarely Perceived. The Dimensions of Intervention. On Management and Buying Time. Biographical Notes and Bibliographies

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

THE AUTHORS

NORBERT HIRSCHHORN and WILLIAM B. GREENOUGH III ("Cholera") are respectively associated with the U.S. Public Health Service and the Johns Hopkins University School of Medicine. Hirschhorn writes: "Born in Vienna, raised and educated at points between Boston and New York. Three months in Surinam as a medical student and three years in Dacca with the U.S. Public Health Service have wedded me to a career in international medicine. My strong bias is that population cannot be controlled until human health improves sufficiently. Outside medicine I enjoy amateur acting and scribbling poetry for my own consumption." Greenough, who was graduated from Amherst College in 1953 and the Harvard Medical School in 1957, spent three years with the Pakistan-SEATO Cholera Research Laboratory in Dacca; he writes that his "avocational experience with small boats on the waterways of New England and New York State was invaluable in establishing river communications between the main-base laboratory in Dacca and the cholera-vaccine field trial area 35 miles away in rural East Pakistan." From 1965 to 1967 Greenough was at the National Heart Institute, where he was coinventor of a continuous-flow centrifuge for the separation of the various formed elements of the blood. Since 1967 he has been at Johns Hopkins, where he is now chief of the infectious diseases division and associate professor of medicine and microbiology in the School of Medicine.

GEORGE F. BASS ("A Byzantine Trading Venture") is associate professor of classical archaeology at the University of Pennsylvania. He obtained his Ph.D. there in 1964 after study at Johns Hopkins University and the American School of Classical Studies in Athens. Much of his work has been in underwater archaeology, although this summer he is beginning a land excavation of a Bronze Age site in southern Italy. The work he describes in his article was sponsored by the American Philosophical Society, Bauer Kompressoren, the Catherwood Foundation, the Corning Museum of Glass, Nixon Griffis, the Littauer Foundation, Mr. and Mrs. James P. Magill, the Main Line Diving Club of Philadelphia, the National Geographic Society, the National Science Foundation, the Office of Naval Research, the Rockefeller Foundation and William Van Alen.

DALE D. HARPSTEAD ("High-Lysine Corn") is professor in and chairman of the Department of Crop and Soil Sciences at Michigan State University. After obtaining his bachelor's and master's degrees at South Dakota State University he joined the department of agronomy there and, during leaves of absence, worked on his Ph.D. at the University of Nebraska. After receiving the degree in 1961 he joined the staff of the Rockefeller Foundation as a maize geneticist stationed in Colombia. He took up his present work in 1969. "As a youth in South Dakota I developed a lasting love for the open spaces," he writes. "I am very concerned about conservation and the wise use of our natural resources." He adds that he enjoys fishing, hiking and home carpentry and occasionally finds time "to chip away at what I call wood carving."

LYNN MARGULIS ("Symbiosis and Evolution") is with the department of biology at Boston University. She writes: "I was born on the University of Chicago campus; most of my education and certainly all my attitudes about learning were obtained at that very exciting institution. Since I left for the University of Wisconsin in 1957 and the University of California at Berkeley (where I received my Ph.D.) in 1960 I have not lifted my nose from the literature of professional biology, except to look through my phase-contrast microscope. Minor exceptions to that general truth are also related to biology: at age 14 I worked on an Israeli commune; one summer I studied native healers in Mexico, and I trained Peace Corps volunteers for biology teaching in Colombia. The major exception to the above general truth involves bearing and raising my four children, mostly boys, all incessantly active."

PALMER DYAL and CURTIS W. PARKIN ("The Magnetism of the Moon") are at the NASA–Ames Research Center in California; Dyal is a research scientist in the Space Physics Branch and Parkin is a postdoctoral associate. Dyal was graduated from Coe College in 1955 and took his Ph.D. at the University of Illinois. He writes that he enjoys canoeing, hiking, trout fishing, gardening and amateur beekeeping and that he also reads extensively on the history of World War II. Parkin was graduated from the University of Texas at El Paso in 1963 and received his master's degree and his Ph.D. from Vanderbilt University. "One of my main interests outside lunar magnetism," he writes, "is music—from rock to Bach—preferably in live concert or quadraphonic sound. I always carry a book, most often a Greek classic or a science-fiction novel. I travel at every opportunity, finding equal pleasure in a European trip and a wild weekend blazing over blacktop roads on my Honda. Other interests include flamenco guitar, classic film festivals and weekend encounter-group marathons."

JEFFREY M. CAMHI ("Flight Orientation in Locusts") is assistant professor of biology at Cornell University. He has been interested in insects since he was an undergraduate at Tufts University, and he began his present line of inquiry into behavioral neurophysiology while he was a graduate student at Harvard University. He writes that he is attracted by "the almost ideal conditions of insects for studying the development of behavior and of the nervous system."

RICHARD C. ATKINSON and RICHARD M. SHIFFRIN ("The Control of Short-Term Memory") are respectively professor of psychology at Stanford University and associate professor of psychology at Indiana University. Atkinson did his undergraduate work at the University of Chicago and received his Ph.D. from Indiana in 1955. His work has ranged from studies of memory and perception to applied problems of learning, particularly teaching reading in primary grades. Shiffrin, who is currently on leave from Indiana to serve as a visiting member of the faculty at Rockefeller University, obtained his bachelor's degree in mathematics from Yale University in 1964 and did graduate work in psychology at Stanford, where he received his Ph.D. in 1968.

LYNN ARTHUR STEEN ("New Models of the Real-Number Line") is associate professor of mathematics at St. Olaf College in Minnesota. During the academic year 1970–1971 he was on leave to work at the Institut Mittag-Leffler in Sweden. Steen has been interested in philosophy, mathematics and physics since his undergraduate days at Luther College in Iowa. In 1965 he received his Ph.D. in mathematics at the Massachusetts Institute of Technology, with a minor in philosophy and mathematical logic. He writes that one of his principal interests is "to communicate some of the excitement of modern mathematics" to people who are not specialists in the field.

The one aspect of college life that seems to remain constant is that each new group of students feels it can do a better job of helping society meet its present and future needs. We think this holds as true today as it ever did.

It seems to us that campus unrest is publicized well out of proportion to the reality of the positive goals of the great majority of students. Students who, above all, want to make a meaningful contribution to our society with their lives.

Of all the reasons we have for supporting colleges, our belief in young people is the most important. So, each year, we continue our aid-to-education program which, this year, touches some 300 colleges and universities.

Many of the young students who benefit from our assistance to colleges and universities will never work for Texaco. But among them will be leaders important to our future. They include a wide range of young people with some very strong beliefs that they can do something worthwhile. All in all, we think we're getting a pretty good return on this investment.

TEXACO

We're working to keep your trust.

We still believe in them. And we're backing that belief.

Established 1845 SCIENTIFIC AMERICAN August 1971 Volume 225 Number 2

Cholera

The disease can readily be treated by the replacement of lost body fluids, but this calls for effective social organization. Knowledge of how the toxin acts may point toward a simpler specific antidote

by Norbert Hirschhorn and William B. Greenough III

In 1849 an explosive epidemic of cholera struck the city of London, killing thousands of its citizens. An astute and studious physician, John Snow (inventor, pioneer in inhalation anesthesia, obstetrician to Queen Victoria), carefully mapped the incidence of the disease and determined that most of the people who died had drunk water from the Broad Street pump. The water, drawn from the Thames River, had the taste and odor of sewage. Snow advised the guardians of the parish to remove the pump handle, and the epidemic quickly waned.

Snow was an objective enough scientist to note that the epidemic had begun to subside before this water source was cut off. Yet he was convinced that the disease must have been caused by some agent in the water associated with human fecal waste. Although the existence of germs had not yet been demonstrated in those days, Snow deduced that the cholera agent was a self-replicating organism, because its effect was not diluted by passage in water.

Snow's well-reasoned conjecture that the epidemic had arisen from the contaminated water, not from "miasmic vapors" or supernatural causes, was soon dramatically confirmed. In 1854 the city was struck by another severe cholera outbreak. In the intervening years one of the two private water companies serving the city had changed its water source, drawing from a clean area of the Thames instead of from the sewage-laden area. Among the users supplied by this company, the Lambeth Company, relatively few persons were affected by the 1854 cholera epidemic, the number of deaths in a 14-week period being 461, or 2.6 per 1,000. In the population served by the other company, Southwark and Vauxhall, which was still drawing from the sewage-laden waters, the deaths from cholera totaled more than 4,000, or 15.3 per 1,000.

The cholera agent had been reintroduced by sailors from the Baltic region. Snow, in his now classic treatise on the two epidemics titled "On the Mode of Communication of Cholera," observed that the separation of the two companies' water sources had set up an "experiment . . . on the grandest scale." Both companies had originally offered their water service to householders competitively, and as a consequence their customers were intermingled throughout the city, each street having some Lambeth users and some Southwark and Vauxhall users. In the 1854 situation, as Snow pointed out, "no fewer than 300,000 people of both sexes, of every age and occupation, and of every rank and station . . . were divided into two groups . . . one group being supplied with water containing the sewage of London, and amongst it whatever might have come from the cholera patients, the other group having water quite free from such impurity." The result of the large-scale "experiment" left no doubt that the sewage carried the cholera agent.

A great deal has since been learned about cholera, but Snow's perceptions still stand as a model of accuracy and erudition. Even today his observations on the nature of the disease sound amazingly modern, as we shall see. We shall review here the current research, which continues to be active, looking toward even more effective control of the disease than has already been achieved.

Although cholera was virtually eliminated from the Western Hemisphere and Europe and confined to certain areas in Asia during recent decades, it has by no means been reduced to a minor threat. There are still massive local outbreaks such as the current one among Bengali refugees from East Pakistan. One such outbreak began to spread in 1958 and is now seeded throughout the Middle East and Africa and in parts of Europe. Arising not from a focus in India but from Celebes Island in Indonesia, and generated by a variant type of the cholera bacterium, that outbreak has already reached the dimensions of a pandemic. Like the previous pandemics, it seems likely to cross the Atlantic, taking hold particularly in the slums and along the rivers of South and Central America.

Cholera is surely one of the most terrifying of all diseases. It strikes suddenly and massively, sweeping through entire families, villages and cities. The symptoms typically begin with a vague feeling of fullness in the abdomen and loss

of appetite. The victim becomes giddy, his hands and feet become cold and clammy, he may vomit and he begins to pass great quantities of liquid stool, at first brown and then the "rice water" stool—opalescent, watery and flecked with mucus—that is characteristic of acute cholera. Excruciating cramps seize the victim's limbs and abdomen. He may quickly go into deep shock. Death, essentially from the effects of dehydration, may come within hours or after several agonizing days. Between half and three-quarters of those acutely stricken will die if they do not receive prompt treatment.

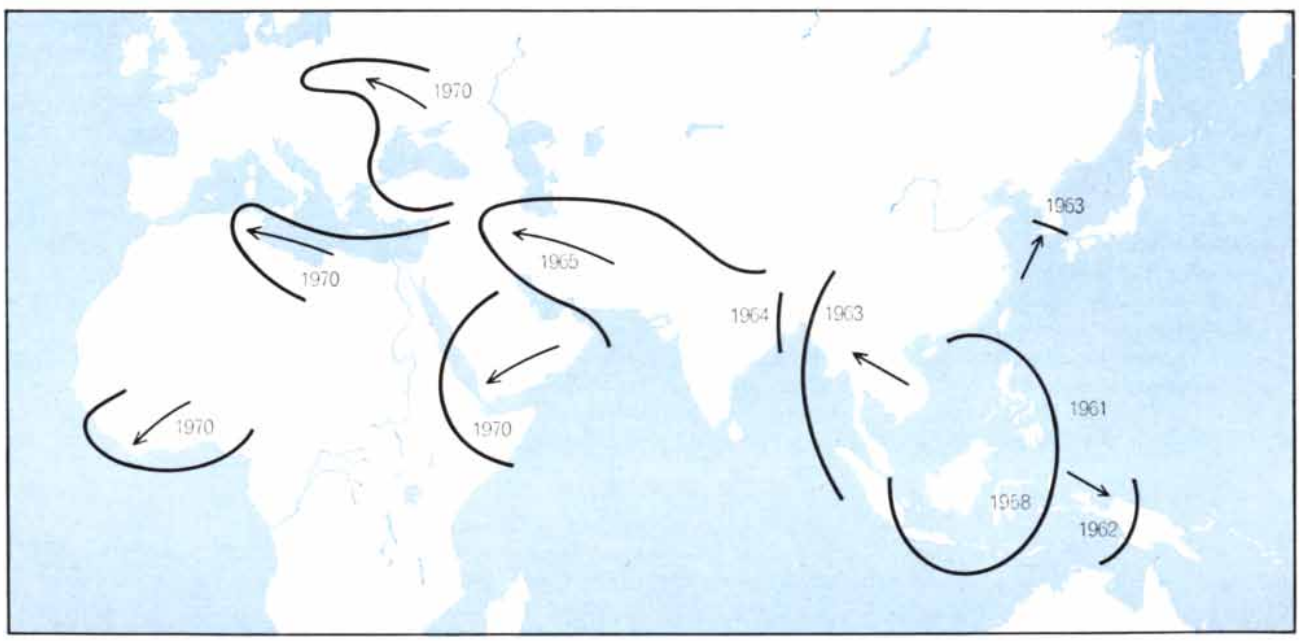
It is no wonder that a cholera outbreak almost invariably throws the population into panic. Many people, including physicians, flee from the stricken area. Bordering countries attempt to stand off the disease by embargoing the movement of people and goods; the epidemic nevertheless spreads inexorably.

Cholera has probably plagued mankind for thousands of years; ancient Sanskrit references give clinical descriptions of an epidemic, cholera-like illness. The disease has certainly been endemic in the Indian subcontinent, particularly along the great river systems, for centuries. The Western world remained relatively free of cholera until the 19th century; between 1817 and 1923, however, it too was visited by a series of six pandemics, each lasting for about a decade or longer. Most technologically advanced countries have been free of the disease since early in this century. This has been attributed to—but may not be solely the result of—sanitation and public health measures. Sanitation, health, wealth and personal cleanliness do not, however, provide a sure barrier against cholera, as was shown recently by an outbreak in Israel. It can occur whenever food or water is contaminated by fecal material containing the bacterium; in Israel vegetables washed in contaminated water were sold in districts with excellent sanitation and spread the disease among well-to-do customers. Moreover, as a disease of man, cholera can be spread in today's world as far and as fast as man can fly.

It is now time to consider the organism that does all this damage. Cholera is caused by *Vibrio cholerae*, a member of the large family of water bacteria that are named vibrios from their darting movements, propelled by a single flagellum. The organism was first identified by the pioneer bacteriologist Robert Koch in his study of cholera in Egypt and India in 1883. A number of investigators have since given intensive study to the cholera vibrio. It is actually a delicate creature, thriving only in a salty, alkaline environment such as exists in tidal estuaries and the human small intestine. (The extraordinarily observant Snow noted that the Southwark and Vauxhall water was saltier than Lambeth's!) Because of its delicacy the cholera vibrio can be cultured only in a special medium that suppresses more vigorous bacteria.

With proper techniques the presence of the vibrio in a fecal sample or a swab can be detected at a patient's bedside within minutes. Under a dark-field microscope the organism's darting movements are easily seen; against the dark background a field of vibrios looks like a speeded-up old-time movie of Times Square at night. When a drop of serum from a rabbit that has been immunized specifically against cholera is placed on the slide, the vibrios clump and all movement ceases; thus the diagnosis of cholera is confirmed. A pocket version of such a microscope has been developed as a tool for identifying the vibrio of cholera and the spirochete of syphilis.

Cholera is not very contagious; the organism generally is transmitted only in food or water and not from person to person, unless one allows one's hands to be soiled by a sick person's evacuations and fails to wash before eating. It has also been found that subclinical infection by the cholera vibrio, in which the host shows no sign of illness, is common. This situation is similar to that of poliomyelitis or viral hepatitis, where the manifest cases of illness represent only the tip of the iceberg, with a much larger number of persons carrying the infection and spreading it unknowingly. In endemic focal regions such as India many people develop a natural immunity to the disease with age, no doubt through repeated subclinical exposures to the vibrio. About half of the victims stricken by cholera are children who have not developed immunity.



CURRENT WORLD OUTBREAK of cholera began in 1958 on Celebes Island in Indonesia. The map, from the World Health Organization, shows the spread of the pandemic year by year since then. The outbreak is considered likely to cross the Atlantic Ocean.

Virtually all patients attacked by cholera can survive if they are promptly and properly treated. The lifesaving treatment is simply replacement of what they are losing in the watery stool. Roughly speaking, all the derangements in cholera are due to the loss of a salty, watery fluid from the intestine. Again we can hark back to the insight of Snow, arrived at in a day when little was known of physiology or biochemistry. Drawing on his own observations and those of contemporary physicians, he wrote:

"The stools and vomited matters in cholera consist of water, containing a small quantity of the salts of the blood, and a very little albuminous substance. The change in the blood is precisely that which the loss by the alimentary canal ought to produce.... If any further proof were wanting... that all of the symptoms attending cholera, except those connected with the alimentary canal, depend simply on the physical alteration of the blood, and not on any cholera poison circulating in the system, it would only be necessary to allude to the effects of a weak saline solution injected into the veins in the stage of collapse. The shrunken skin becomes filled out, and loses its coldness and lividity... the patient is able to sit up, and for a time seems well."

Snow's remark about the effectiveness of treatment with intravenous injections was excessively cautious; he put his finger, however, on the jugular of the cholera medical problem. In our century a number of investigators, notably Robert A. Phillips, Raymond H. Watten, Craig K. Wallace and their associates working in a research unit of the U.S. Navy in Taiwan, explored in detail the losses "by the alimentary canal" in cholera and developed specific principles for treatment of the attack.

The volume of the loss of body fluid is in itself devastating. Cholera patients have been known to lose fluid at the rate of more than a liter an hour for many hours during their diarrhea. On admission to a hospital they commonly show a loss of from 5 to 10 percent of their body weight, which is indeed a massive loss considering that this amounts to about a quarter to half of the normal total weight of the body fluid outside the cells. The patient shows the outward signs of the effects of fluid loss: a faint or absent pulse, collapsed veins, sunken eyes, loss of skin elasticity, diminished or absent urine output and, in very severe cases, shock with loss of consciousness, due to the lack of adequate blood circulation to the vital organs.

The investigators in the Taiwan re-

IMMUNIZATION against cholera is provided at an "inoculation barrier" set up by the Indian government near a pilgrimage center. The high concentration of population and the lack of adequate sanitary facilities associated with pilgrimages (and with the current flood of Bengali refugees) often lead to an outbreak of the disease, which is endemic in India. The killed-whole-cell vaccines are at best only partially effective, and only for about three months at that. The potency of commercial vaccines varies; some of them give no protection.

CHOLERA PATIENT is treated by intravenous injection of fluid to make up for the catastrophic fluid loss characteristic of the disease. The severe dehydration reduces the blood volume sharply, leading to deep shock and unconsciousness. The outward signs include lack of a pulse and of urine flow and the flaccid skin and sunken eyes seen in this patient.

search unit found that the blood of a typical patient contains higher than normal concentrations of protein and blood cells, a nearly normal concentration of sodium in the serum and a marked decrease of alkali. These facts indicate, and analysis of the watery stool confirms, that little protein is lost but a considerable quantity of salt and bicarbonate is carried off in the stool.

Once the important losses had been identified, the necessary treatment became obvious: inject into the patient a solution that matches in amount and composition what is lost in the stool. This is supplied by continuous intravenous injection. The volume of excretion of stool is so high that over a treatment period of from four to seven days the total infusion for replacement may amount to twice the weight of the patient's body or more. To keep track of the patient's loss of fluid, and to provide a hygienic method of stool disposal, Watten of the Taiwan team designed a special "cholera cot." The cot has a hole under the patient's buttocks, and a plastic sheet with a sleeve guides the stool into a calibrated bucket below.

The treatment produces a dramatic recovery. Within minutes the patient regains warmth and alertness. If hydration is maintained properly, he can soon sit up on the cot and even take nourishment. In four to seven days the diarrhea ceases and the patient is cured and can go home.

The composition of the replacement solution is also important. If the bicarbonate is not replaced, particularly in cases where volume replacement has also been inadequate, there is an increase in the concentration of hydrogen ions in the blood, resulting in acidosis. A patient whose blood *p*H falls below 7.0 is perilously close to death. Acidosis weakens the pumping of the heart and causes a constriction of the blood vessels that tends to shunt much of the remaining blood volume into the lungs. This leads to congestion of the lungs and sets the stage for either pneumonia or pulmonary edema if an adequate degree of rehydration is attempted before the acidosis is corrected.

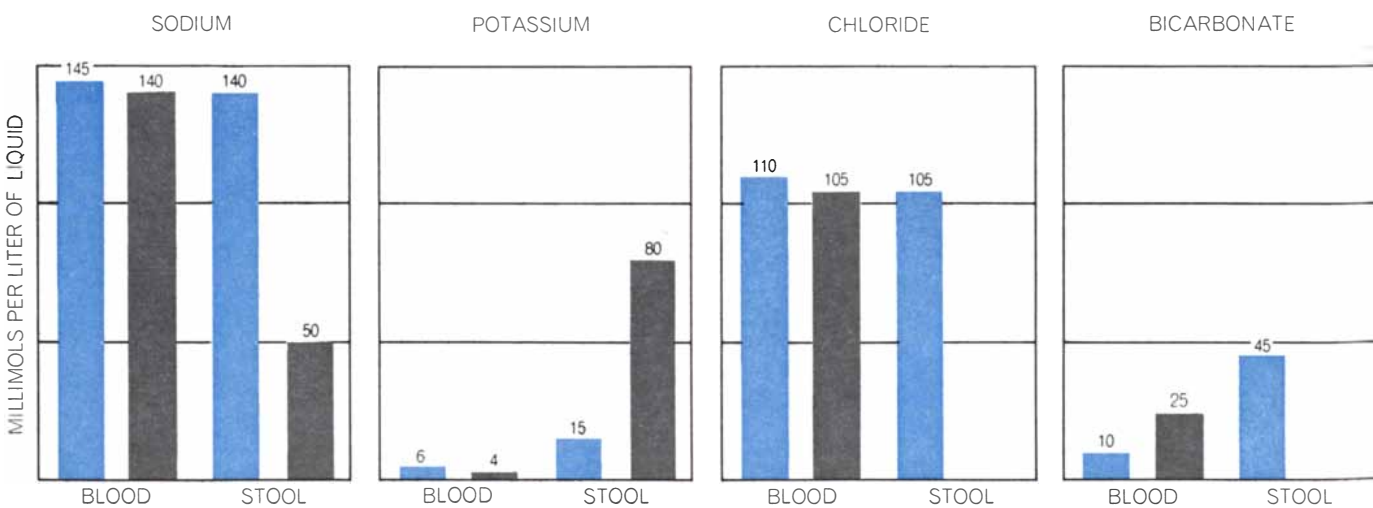
In small children, or even in adults, when diarrhea is prolonged, the loss of another salt, potassium, can lead to disturbances of the heart rhythm and even to heart stoppage. Potassium depletion also causes profound weakness and possibly paralysis. Most replacement solutions therefore include potassium.

Where the treatment is available, the mortality from cholera can be reduced to less than 1 percent. Unfortunately it is not yet available in most of the cholera-ridden areas of the world. It requires large quantities of sterile intravenous fluids, skilled personnel and special facilities. The World Health Organization is making superb efforts to train personnel and supply the necessary material, but its resources are too small to bring the required protection to many of the areas at risk from cholera.

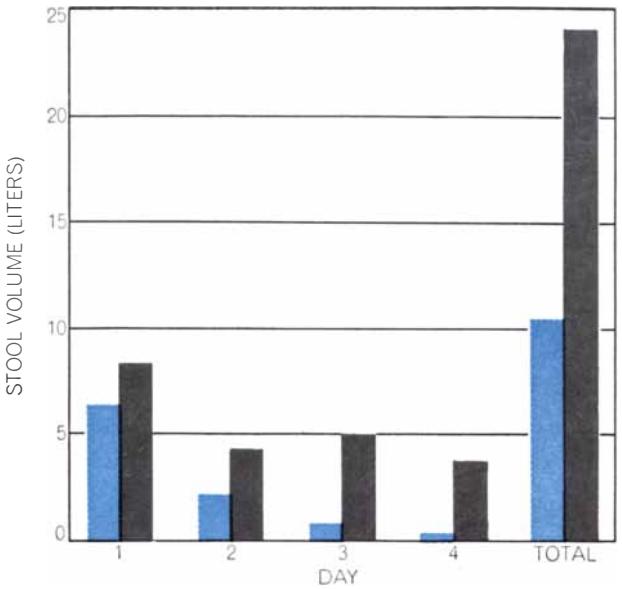
A new laboratory development is helping to relieve the central problem: the difficulty and expense of replacing the lost body fluid by intravenous injection of sterile fluids. It has been found that simple sugars such as glucose and certain amino acids such as glycine markedly increase the transport of sodium and with it water across the intestinal wall. This has opened the way to supplying the necessary fluids by mouth instead of by injection into the bloodstream. An oral solution approximating the composition of cholera stool and with glucose as the carrier has been developed and tested by David Nalin, Richard A. Cash and one of us (Hirschhorn), working in Dacca in East Pakistan, and by Nathaniel F. Pierce, R. B. Sack and John G. Banwell in Calcutta, and by others. The patient is first rehydrated by intravenous injection, if necessary, and is then fed the oral solution at the rate at which fluid is lost by diarrhea. With a thin stomach tube one can continue the feeding while the patient is asleep.

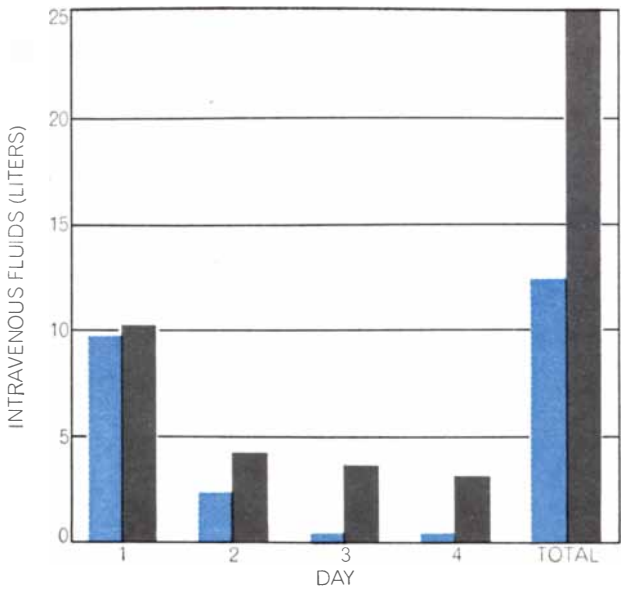
The new method can reduce the requirement for intravenous injection to as little as 20 percent of the total fluid needed for the cure. More significant, since local water sources can be used and the salts and glucose are readily available in the market, the expense is sharply reduced; in a rural center in East Pakistan, for example, the cost of the fluids for the full treatment was cut from $42 per patient to 63 cents. In epidemic situations, where intravenous fluids are often not available, early treatment with the right oral solution could save many lives. In its simplest form oral treatment could be initiated by letting a patient's relatives have a quantity of the solution with instructions to give at least one or two quarts an hour until good urine output has been achieved. Such an approach would provide a first line of defense where medical facilities are overwhelmed, as in the Bengali refugee camps.

A vibrio-killing antibiotic such as tetracycline can also help, as was demonstrated in Dacca by a group including John Lindenbaum and one of us (Greenough) and in Calcutta by Charles C. J.



COMPOSITION of the blood and stool in severe cholera provides guidance in making up the replacement fluids and also yields clues to the mechanism of the disease. Colored bars give average values in cholera, gray bars normal values, where pertinent (some normal





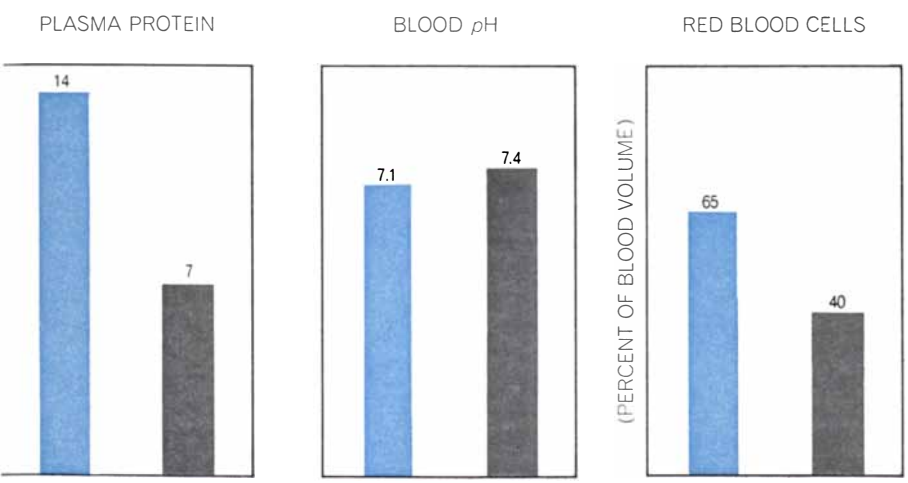
MAGNITUDE OF FLUID LOSS is demonstrated by data on the stool volume (almost completely liquid) and replacement intravenous fluids in a series of patients treated in Calcutta by Charles C. J. Carpenter and his colleagues. In patients given fluids but no antibiotics the loss was still large after four days (*black bars*). An antibiotic reduced the duration of diarrhea significantly (*color*).

Carpenter and his colleagues. The antibiotic cuts the duration of the diarrhea in half, thus saving time, money and distress for the patient. It is not, however, a cure for the disease; only the replacement of the patient's fluid losses can restore him to health.

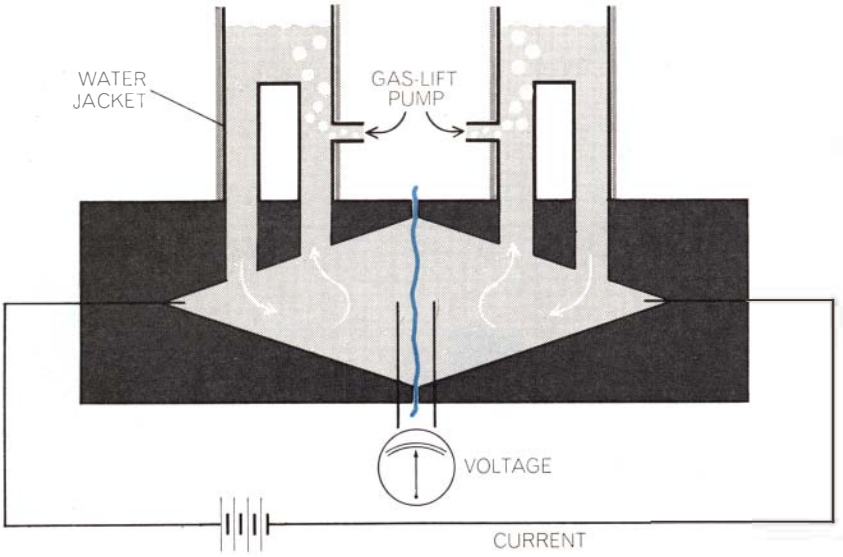
The progress that has been made in treating cholera should not give us an overconfident sense of security in the Western world. In richly fed populations the high incidence of atherosclerosis (causing narrowing of the arteries) creates a hazardous setting for any dehydrating disease. The sudden drop in blood volume and pressure in cholera, reducing blood flow through narrowed vessels to the heart and brain, could produce an epidemic of heart attacks and cerebral strokes. Hence there is good reason to pursue further research into the basic processes of cholera in an effort to find means of forestalling the disease or suppressing it as quickly as it strikes. Actually the continuing research has turned up biological findings that go considerably beyond the disease in interest and significance.

The most fundamental question about cholera demanding an answer has been: How does the cholera vibrio bring about the body's massive outpouring of fluid? Once again Snow's remarkable intuition is shown in his framing of this question more than a century ago: "It would seem that the cholera poison... withdraws fluid from the blood circulating in the capillaries, by a power analogous to that by which the epithelial cells of the various organs abstract the different secretions in the healthy body." The key word here is his reference to a "power" driving the fluid from the blood vessels into the intestine. It has now been established by a number of investigators that the vibrio is found all up and down the intestine, swimming in the lumen and clinging to the lining of the intestinal wall. Intubation experiments, in which fluid is sampled at various levels in the intestine, show that the outpouring of fluid comes only from the small intestine, not from the stomach or the colon. Other studies show that the intestinal wall remains intact and that the fluid in the intestine is nearly normal in composition (although not in amount!). All these findings, confirmed in animal studies by Sack and Carpenter at the Johns Hopkins University School of Medicine, suggest that power, or energy, is somehow supplied for driving the excessive flow of fluid from the blood into the intestine. One might suppose that the cholera vibrio simply makes the intervening membranes "leaky" in some way, but experiments have shown that the permeability of the intestinal membranes is not increased. It turns out that the vibrio does in fact trigger a process that stimulates secretion of the fluid.

Several independent lines of inquiry, eventually drawn together, led to the elucidation of this process. The findings were based on the work of investigators



values depend on diet). Sodium and chloride in blood are elevated by loss of water through sweating and labored breathing. The rise in blood potassium is due to acidosis.



EXPERIMENTAL CHAMBER for testing the passage of ions through intestinal tissue was employed by Michael Field of the Harvard Medical School. Electrical gradients detected by the voltmeter are neutralized by a current. Substances affecting the tissue are added to the fluid and changes in active transport of ions by the tissue are measured by sampling the fluid.

in Calcutta, Bombay and Dacca, at the University of Oxford, the Harvard Medical School, Johns Hopkins and the University of Minnesota.

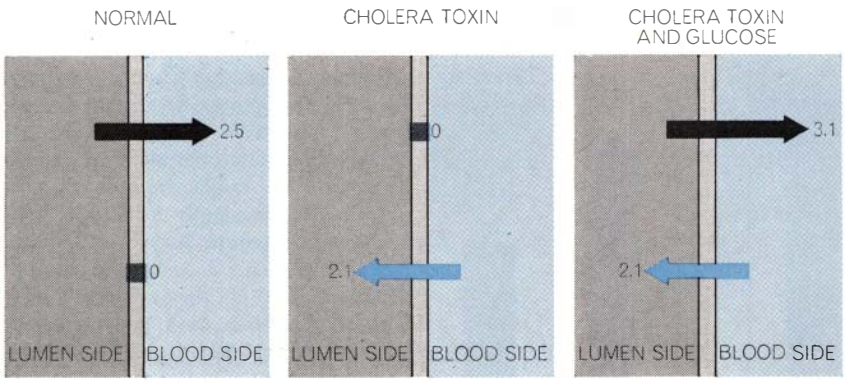
Michael Field of the Harvard Medical School had for several years been studying absorption and secretion by the small intestine of the rabbit. Using a preparation of intestinal epithelium in which he could neutralize any physical, electrical or chemical gradients acting across the tissue, he traced the movement of salt ions from the lumen side of the intestinal wall toward the blood and vice versa [*see illustration above*]. He found that normally the sodium ion of salt was moved by the membrane to the blood side, corresponding to the absorption of salt and water by the blood from the intestine in an animal. However, when Field added to the tissue preparation the drug theophylline, a component of tea long used as a heart stimulant, the flow was reversed: chloride ion moved toward the intestinal lumen side and sodium was no longer absorbed, corresponding to secretion from the blood of salt and water.

Theophylline has a highly specific effect: it fosters the body's accumulation of the substance known as cyclic AMP (a form of adenosine monophosphate) by inhibiting an enzyme that breaks down this substance. Cyclic AMP is called a "second messenger," because it passes on orders from various hormones to the cells. Field proceeded to test the effect of cyclic AMP on his rabbit intestinal tissue and found that it had precisely the same effect there as theophylline: it reversed the flow of salt ion from the normal absorption direction to the secretion direction.

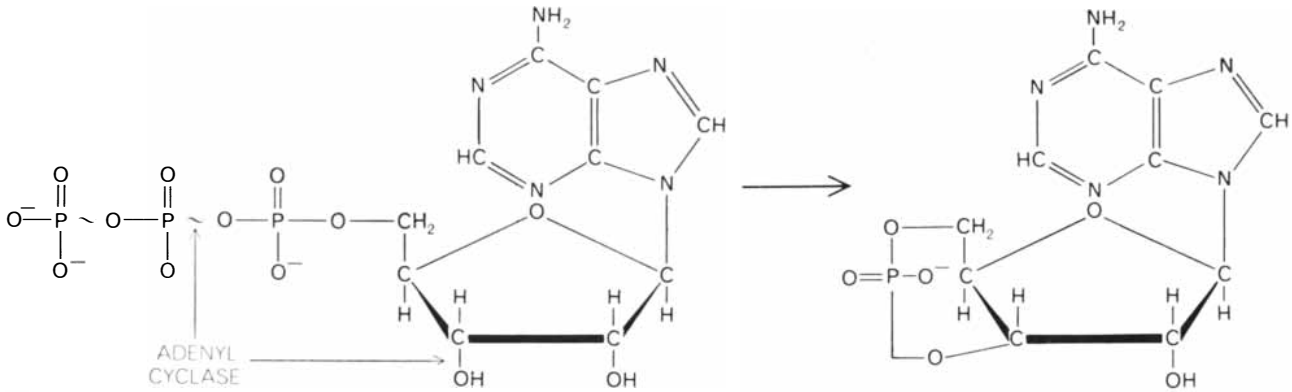
One of us (Hirschhorn), with Howard S. Frazier of Harvard, later performed another experiment with theophylline. Examining the reactions of rabbit intestinal cells by means of microelectrodes inserted into the cells, we found that the drug produced an electrical effect within the cells that was consistent with chloride-ion secretion. We observed this effect mainly in the basal cells, which are thought to be the main sources of secretion by the small intestine. Presumably the cyclic AMP was responsible for the cellular secretion we were observing. Here, then, was a secretory process tailor-made to produce fluid loss from the intestine.

We switch now to another main thread of the story. More than a decade ago three Indian investigators, Nirmal K. Dutta in Bombay and S. N. De in Calcutta, discovered that not only the cholera vibrio itself but also the fluid in which the organism had been cultured could cause copious secretion of salt and water from rabbits' intestines. Findings by several investigators, including William Burrows of the University of Chicago and George F. Grady of Tufts University, implicated a specific toxic substance produced by the vibrio. After years of work independent groups headed by Richard A. Finkelstein of the Southwestern Medical School, Stephen H. Richardson of the Bowman-Gray School of Medicine and W. H. Moseley of the Pakistan-SEATO Cholera Research Laboratory in Dacca isolated a purified toxin that, in amounts as small as a ten-millionth of a gram, could produce the fluid loss characteristic of cholera from rabbit intestinal tissue.

With the discovery of the effect of cyclic AMP the riddle began to unfold. Could it be that the vibrio toxin issued a signal in response to which cyclic AMP, as a second messenger, opened the gates for a flood of secretion from the intestinal cells? In 1968 the authors of this article gave a seminar on cholera at Harvard that Field attended, and Greenough and Field formed a Harvard-Hopkins cooperative research project. The first experiments proved that the vibrio product that produced cholera in animals did indeed set off the same secretory process that cyclic AMP did. From the experiments with rabbit tissue Quais al Awqati and Greenough went on to tests of surgical specimens of tissue from the small intestine of man, and this tissue also responded to either cholera



NET ION FLOW across intestinal tissue is given in microequivalents per square centimeter per hour. The normal flow of sodium ion (*top*) and chloride ion (*bottom*) is reversed (*center*) if cholera toxin is added to fluid. Theophylline or cyclic adenosine monophosphate (cyclic AMP) has the same effect. Adding glucose on lumen side helps restore balance (*right*).



SPECIFIC EFFECT of the cholera toxin is exerted through the enzyme adenyl cyclase. The enzyme converts the cell's primary energy-carrier, adenosine triphosphate (ATP), into cyclic AMP; the higher AMP level causes the intestine wall to secrete chloride (and with it water) into the intestine. The coincident decrease in the absorption of sodium (and with it water) enhances the fluid loss.

toxin or cyclic AMP with secretion of salt and water.

The next step in linking cyclic AMP to the toxin was a foray into this question: Would the cholera toxin have an effect on other tissues—that is, tissues of organs other than the intestine—whose cells respond to cyclic AMP? The toxin is not absorbed from the gut into the bloodstream, but it was no difficult matter to test it in animals and on isolated tissue specimens. The toxin did have a general potency when injected into dogs; Pierce has shown that the effects mimic what might be expected from hormonal derangements. In isolated tissues very small amounts of toxin, less than a billionth of a gram, caused fat cells to release their stored lipids, liver cells and blood platelets to accelerate the breakdown of their glycogen and human white blood cells to refrain from releasing histamine. Since these responses are known to be mediated by cyclic AMP, it seemed clear that the toxin must stimulate production of the second messenger. Final proof that it does so came simultaneously from three separate investigations, conducted by Field and Daniel V. Kimberg at Harvard, by Geoffrey W. G. Sharp, also at Harvard, and by David E. Schafer at Minnesota. They all found that after intestinal tissue has been exposed to the cholera toxin it does indeed have heightened levels of cyclic AMP, produced through increased activity by the enzyme adenyl cyclase, which converts adenosine triphosphate (ATP), the cells' main source of stored energy, into cyclic AMP [*see illustration above*].

As is so often true in medical science, the intensive investigation of a particular disease has once more in this case provided new insights into normal physiological and biochemical processes. The vibrio cholera has revealed that the small intestine possess a powerful capacity for secretion of fluids, a capacity for removing more than 100 liters of liquid from the human body in less than a week. How is this capacity regulated in the healthy body? We must suppose there are normal mechanisms operating to regulate the absorption and secretion of fluids by way of the intestine. By analogy with the regulation of other tissues, we can assume that the generation of cyclic AMP in the intestine is normally stimulated by hormones that have not yet been identified. One possibility is that this function may be performed by prostaglandins: substances that were originally isolated from the prostate gland and that are known to increase levels of cyclic AMP in many tissues. In the intestine of the dog large doses of these hormones produce a cholera-like secretion of salt and water.

With regard to the cholera toxin, one naturally wonders how it is that the body cells "recognize" and respond to this foreign protein. There are several possible explanations, each of which raises interesting questions. It may be that this substance made by the bacterial cell is similar to a hormone of man, in which case it would appear that hormones have a more ancient origin than has been supposed, and one is led to speculate on what function such a substance would serve for a bacterium. Other possible explanations of how the toxin takes effect are that it may somehow stimulate the adenyl cyclase enzyme without needing to attach itself to the receptor site for a hormone, or that it may induce secretion of a hormone that normally stimulates the enzyme, or that the toxin may itself become an active adenyl cyclase when it is attached to cells. In the last case the uncontrolled outpouring of fluid from the intestine could readily be understood: with a substance produced by a foreign invader acting as the stimulator for the production of cyclic AMP, there would be no "feedback" mechanism to turn it off when too much was produced.

The newly gained knowledge about the process causing the loss of fluid in cholera has, of course, considerable medical significance. Not surprisingly, it has now been discovered that a similar mechanism is responsible for the severe diarrhea that is produced by some other infections. *Escherichia coli* in particular also has a toxin that stimulates secretion of salt and water from the intestine. Two investigators in Dacca, Lincoln Chen and George T. Curlin, demonstrated that the toxin of an *E. coli* found in a patient with severe diarrhea stimulated the adenyl cyclase in fat cells. The finding that the cholera toxin represents a fairly common mechanism for the production of diarrhea may lead to a new approach for dealing with the dangerous volume-depleting diarrheas caused by a number of organisms. These other diarrhea-producing infections actually cause more disease and death than cholera does, particularly among children, and they may be responsible for a significant proportion of the severe malnutrition following acute diarrhea that is seen in impoverished areas of the world.

As more is learned about how such toxins exert their effects, it should become possible to find drugs that will act as specific antidotes. A drug might seize and deactivate the toxin (as an antibody does) before it arrives at the cell surface, or it might occupy the receptor sites on cells the toxin could attack, or it might directly inhibit the stimulation of adenyl cyclase. As an answer to cholera it may come about someday that at the first sign of the disease a poor river fisherman in Asia or the accidentally exposed population of a modern city will be able simply to take a pill and go back to work. Snow would have appreciated that.

A Byzantine Trading Venture

Early in the seventh century a small cargo ship went down in the eastern Mediterranean. Four seasons of underwater investigation have reconstructed the vessel's last voyage

by George F. Bass

In A.D. 625 or 626 a Greek merchant vessel of 40 tons' burden was sailing south along the coast of what is now Turkey when it struck a reef some 20 miles from Halicarnassus and sank in 120 feet of water. In 1958 the wreck was discovered during a diving reconnaissance by Peter Throckmorton, and from 1961 to 1964 the underwater site was excavated by my group from the University Museum of the University of Pennsylvania. Analysis of our findings, which is now reaching completion, provides a vivid picture of a trading venture in the eastern Mediterranean during the days of the Byzantine Empire. This is the story of the ship and its last voyage.

Ten years or so before the ship went down a cypress keel had been laid in a shipyard probably located in the Aegean, in the Black Sea or in the waters between. On the keel, which was about 40 feet long, was mounted a high, curved sternpost of the same wood. (The ship's stem was probably also made of cypress, but that part of the vessel was not preserved.) Once the spine of the hull had been completed the shipwrights went on to construct the sides.

They did not, as we would today, first build a complete skeleton by adding ribs to the spine and then cover the frame with planking. Neither did they follow the practice, customary in earlier centuries, of stiffening the hull by fastening its planks edge to edge with a series of mortise-and-tenon joints no more than four inches apart along the full length of each strake, or strip of planking. Seventh-century construction fell somewhere between these two methods; it represents the continuation of a trend, starting two or three centuries earlier, that cut down the investment in labor required to build hulls in the earlier Greco-Roman style.

DIVING ARCHAEOLOGISTS from the University Museum of the University of Pennsylvania, working at a depth of 120 feet off the small Turkish coastal island of Yassi Ada, guide a wire basket to the surface from the site of the seventh-century Byzantine shipwreck (*opposite page*). The basket contains timbers from the hull of the vessel.

Selecting pine planks, the shipwrights cut mortises in them every three feet or so and fastened the planks together edge to edge by inserting loosely fitted tenons made of oak. The hull was built up from the keel in this fashion, one strake at a time, until the waterline was reached. The builders then scored the inside of the planks to mark where ribs, made of elm, should be placed. The ribs, which rose from the keel to well above the waterline, were secured in place with iron nails driven through the pine strakes from the outside. Four pairs of heavier strakes were then run along each side of the ship; three pairs were nailed in place and the fourth pair was bolted. The uppermost pair ran along the top of the ribs. The spaces between the heavier strakes were filled by nailing additional planks to the ribs without mortise-and-tenon joining.

We know that this kind of compromise with earlier Greco-Roman practice was on the rise because we have also studied a sunken vessel that is some 250 years older than the seventh-century wreck. The hull of the older ship was made with carefully mortised planking; the mortising continued at least up to deck level and possibly all the way to the gunwale. The planks were fixed to the ribs with treenails—long wooden dowels—rather than with iron nails, and smaller dowels held each tenon securely in place. Even by this time, however, the mortise-and-tenon joints were spaced about seven inches apart, or considerably farther apart than before. The trend was destined to continue. The practice of fastening the planks to the ribs with iron nails and omitting mortise-and-tenon joints, found only above the waterline in the seventh-century ship, advanced until the strength of a hull came to depend exclusively on the bonds between the ribs and the planking, as is true today, and mortise-and-tenon hull-stiffening disappeared.

Once the hull was entirely planked, deck beams were inserted across the width of the ship. They were supported at the ends by short *L*-shaped timbers and elsewhere probably by posts [*see top illustration on page 26*]. The ship was now completely decked except for the galley area at the stern and a hatch aft of the single, centrally located mast. There may also have been a smaller hatch forward, but like the stem timber this part of the vessel was not preserved. Deck beams near the stern projected beyond the hull on each side to form a pair of rectangular structures where the steering oars were mounted.

Frederick van Doorninck, Jr., of the University of California at Davis has plotted the position and angle of every nail hole and scored mark visible on the fragments of ship's hull that were salvaged. Largely on the basis of this evidence he has produced a restoration of the ship [*see illustration on next two pages*]. He was helped to visualize some of the ship's less certain features by the work of a gifted amateur marine architect, J. Richard Steffy, who has constructed a series of scale models on van Doorninck's data.

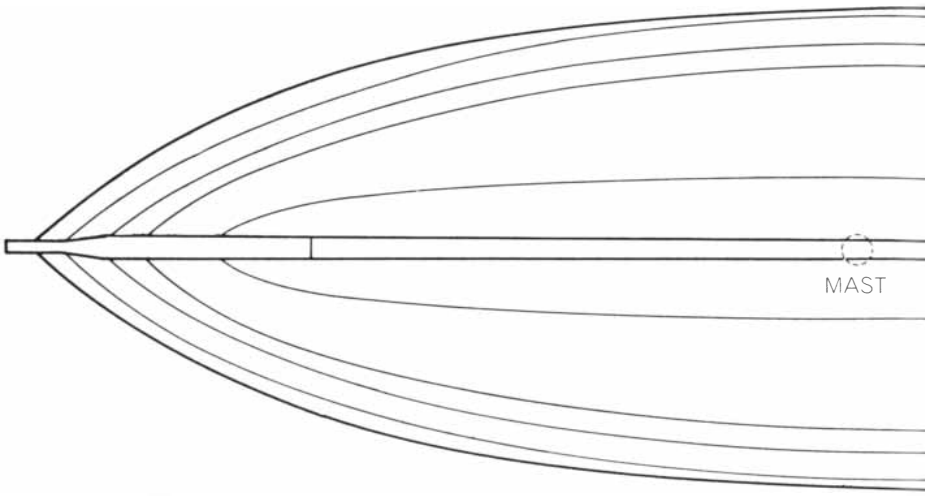
Van Doorninck calculates that the

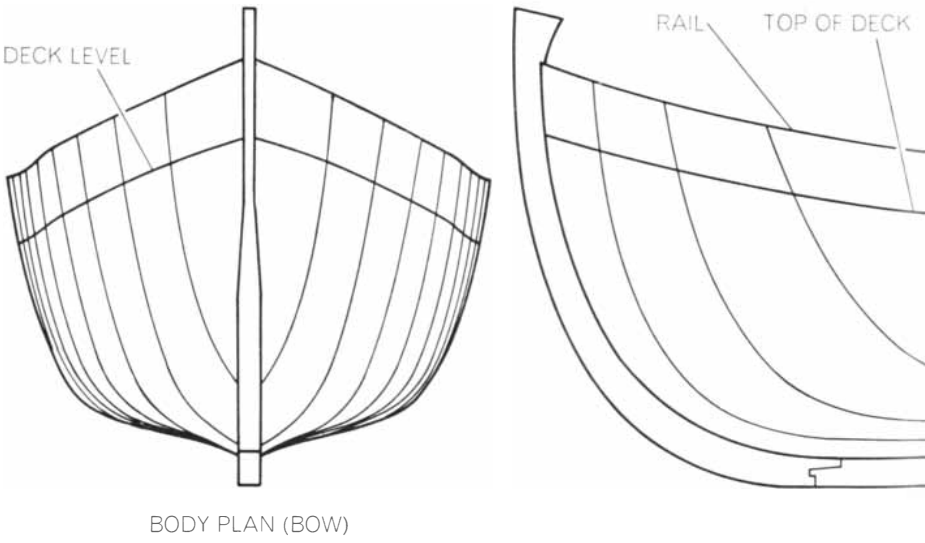
ship was some 63 feet long, with a streamlined hull and a beam of only 17 feet. This is a length-to-width ratio of 3.6 to one, which is quite slender for a cargo vessel. The fourth-century vessel mentioned above had been beamier; its ratio was three to one. It is interesting to note in this connection that several years before either wreck was known a historian, Robert S. Lopez of Yale University, suggested the development of slimmer and swifter merchant vessels in the seventh century because of the need at that time "to dodge or outrun the hostile ships that lurked along every route." Lopez also believes the rise of the independent shipowner in the seventh century must have led to the building of small, single-deck vessels in place of the great cargo ships characteristic of the imperial Roman merchant fleet. Our seventh-century vessel seems to bear out Lopez' contentions. A ship of twice its burden was considered small in Greco-Roman times; many of the vessels plying between Alexandria and Rome carried more than 1,200 tons.

When the slim merchantman was ready for launching, its owner must have visited a ship chandler to buy the gear needed to outfit the vessel. A site recently excavated at the ancient Black Sea port of Tomis—Constanza in modern Romania—has shown us what such a shop must have been like early in the Christian Era. Stacked against a wall in a large vaulted chamber at Tomis were iron anchors with removable anchor stocks, almost identical with the ones found in the wreckage of our seventh-century ship. The owner must have obtained at least 11 such anchors from the chandler, since this is the number we recovered. He also bought a grapnel, probably for use with the ship's boat.

Why were there so many anchors? There are at least two possible explanations. First, the theory and practice of anchoring with the help of heavy anchor chain was unknown to Byzantine seamen, who did not realize that an anchor alone, without the added weight represented by a length of stout chain, is not an efficient bottom-holding device. With only rope for their anchors the mariners probably lost anchors frequently, and they may have been accustomed to carrying spares. The second possible explanation is that the ship was required by statute to carry a specified number of anchors. Such statutes are known to have been in force a few centuries later.

I mentioned that we recovered 11 anchors. What we actually brought to the surface were 11 featureless concretions; the iron that had been at the core of each concretion had corroded away hundreds of years before. We found it possible, however, to cut into the hollow concretions with a jeweler's saw and were thus able to make exact casts of nine of the 11 anchors. Working with the anchor casts, van Doorninck was able to calculate the weight of the originals. Six of the anchors were small, each weighing about 250 Roman pounds (327.45 grams per pound); each of the three larger ones weighed about 450 pounds. The uniformity among the two



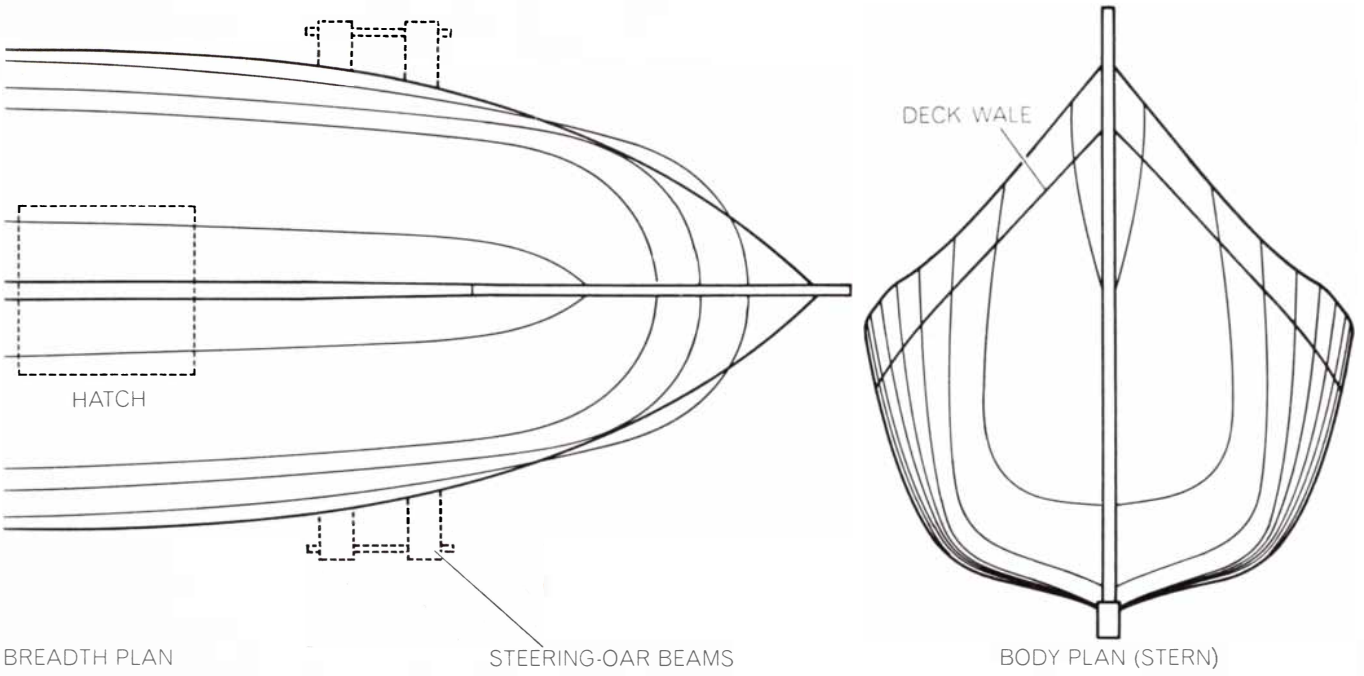


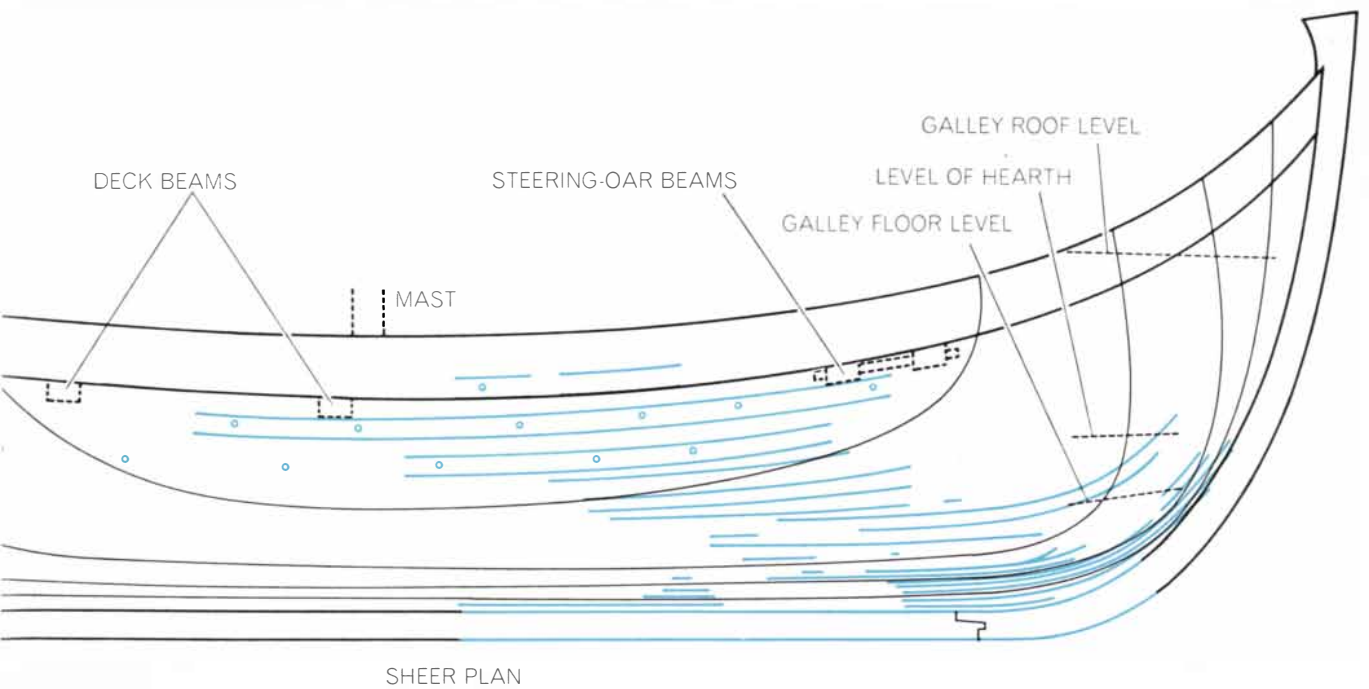
APPEARANCE OF SHIP wrecked at Yassi Ada was reconstructed after detailed study of the wood fragments (*color*) that still remained after more than 1,300 years of submersion. The study was conducted by Frederick van Doorninck, Jr., of the University of California

kinds suggests that in the seventh century, as in later times, there were regulations requiring vessels of various classes to carry a minimum number of anchors of specified sizes.

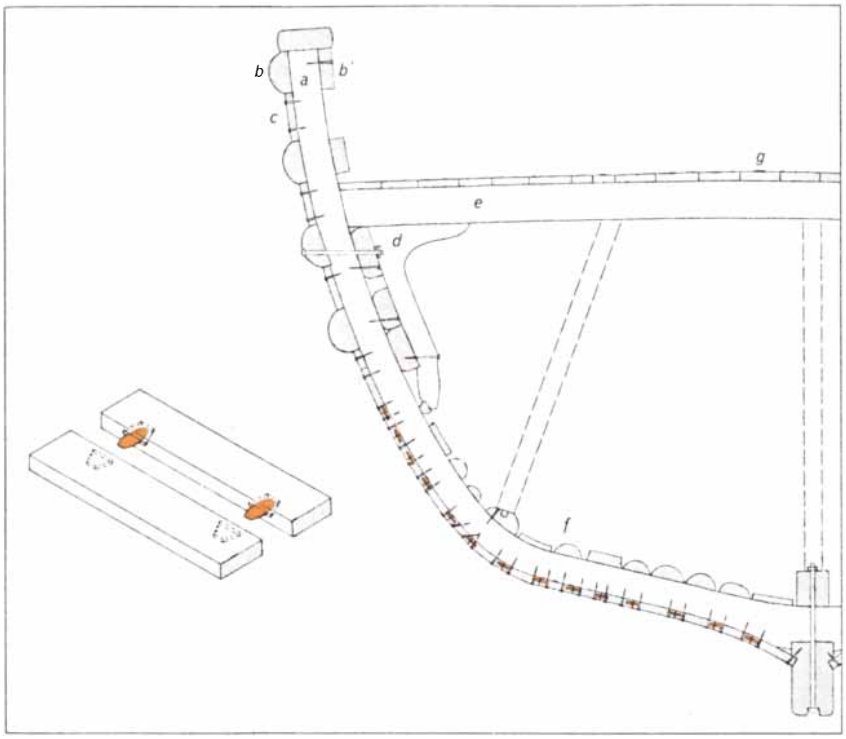
The chandlery at Tomis offered a variety of resins for sale. So, evidently, did the one patronized by our owner. His ship had already been coated with resin below the waterline, inside and out. Now he obtained a smaller amount to be melted on board in a cheap cooking pot and applied as needed to seal the pores of clay pitchers and other containers for liquids. The chandlery at Tomis also offered pigments for paint and jars full of iron nails. We do not know whether or not the owner painted his ship, although literary accounts of earlier and later times speak of many ships as being brightly colored. He did buy several bags of nails, probably with shipboard repairs in mind.

Literary accounts of the period men-

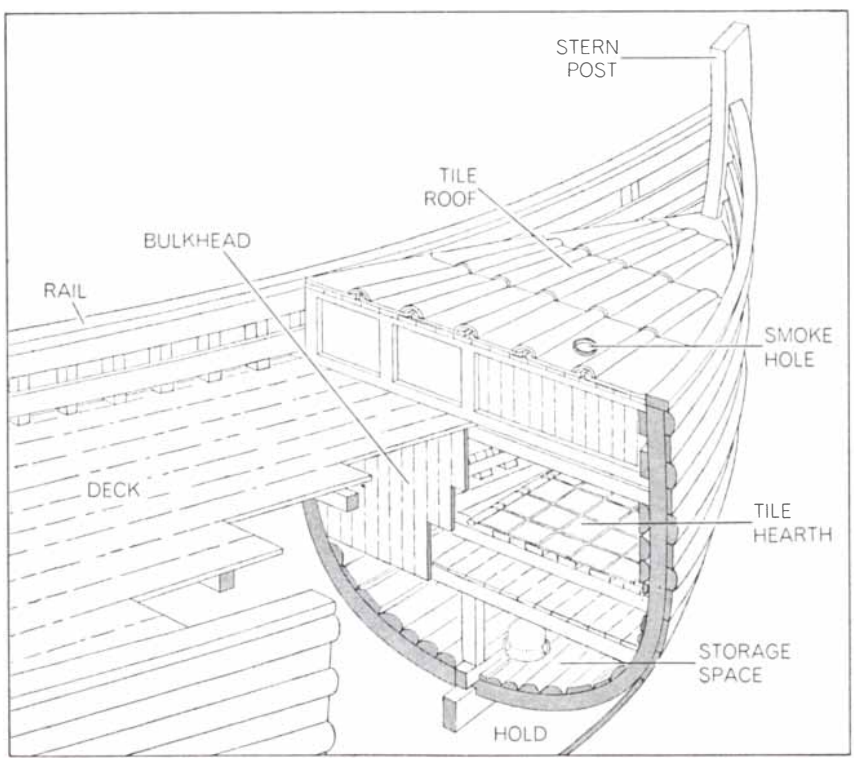




at Davis. The pair of beams projecting outboard just forward of the galley area (*broken lines*) supported the ship's two steering oars. Between the oar beams and the vessel's single mast a deck hatch gave access to the cargo hold; there may have been another hatch forward, but no wood from this part of the ship has survived. The length of the ship at the waterline was some 55 feet.



HULL WAS BUILT by laying planks upward from the keel. The planks were fastened edge to edge with mortise-and-tenon joints (*color*) spaced as far apart as three feet (*detail at left*). When the planking reached waterline height, the ship's ribs (*a*) were inserted and fixed to the hull planking with nails. Four pairs of heavier timbers (*b and b'*) were then laid along each side of the ship, and the exterior was completed with more planking (*c*). *L*-shaped timbers (*d*) were set inside to support deck beams (*e*), which may have needed added bracing (*broken lines*). Hold stringers (*f*) and deck planks (*g*) finished the job.



GALLEY AREA occupied the extreme stern of the vessel and was roofed with flat and curved tiles. The tile hearth was located to port. Its form remains conjectural. One of several arrangements (*not shown*) may have used an iron grid to brace pots against ship's motion.

tion skins for covering deck cargo in foul weather, as well as sails and cordage, and all of these the owner may have got at the chandlery. We are certain that he bought 24 oil lamps, because we found them in the wreck. It seems probable that this was another chandlery purchase; in the shop at Tomis were found not only lamps but also the molds for producing them.

With the new ship almost ready for sea it would have been time to assemble a crew. We need not depend entirely on our imagination in reconstructing the roster; a document of the seventh century or slightly later, known as the Rhodian sea law, lists a regular ship's company of the day and assigns to each member his proper share in the profits of a successful venture. First is the *naukleros,* the shipowner or the captain or both, who receives two shares. Next are the *kybernetes,* the helmsman; the *proreus,* the "prow officer"; the *naupegos,* the ship's carpenter, and the *karabites,* the boatswain. Each receives a share and a half. This completes the list of officers; the *nautai,* or seamen, who follow on the roster receive a single share each. Last is the lowliest member of the crew, the *parascharites.* The approximate meaning of the word is "gut-slitter," and in this context it may mean cook. At any rate, his stake in the voyage is only half a share.

We know that the captain of the vessel was named Georgios. He was probably also the shipowner, or perhaps a part owner. The finest scale on the ship, a bronze balance of the kind called a steelyard beam, has his name punched in it in Greek letters to read *Georgiou Presbyterou Nauklerou.* That is to say "Georgios Elder (or Senior), Owner/Sea Captain." The inscription presents problems in spite of its simplicity. Christianity was a strong element in every aspect of Byzantine life; was Captain Georgios a *presbyteros,* an elder of the church? Or does the word have some other meaning here?

Several passages in the Rhodian sea law indicate that sometimes there was more than one captain aboard a vessel. How would one have distinguished between two captains? Should Georgios' title perhaps be read as "Georgios, Senior Sea Captain"? At the same time that we were excavating the wreck, Martin Harrison of the University of Newcastle discovered an inscribed baptismal basin in Turkey (in a sixth-century church in Lycia). The inscription reads: *Nicholas Naukleros Mesatos. Mesatos* is derived

from *mesos,* meaning middle, and Harrison has tentatively suggested that the term might mean something such as "middling sea captain," that is, an officer junior to a senior sea captain. No one, Harrison included, accepts this suggestion completely. In its support, however, it should be noted that *mesonautai* are mentioned in the literature of the sixth century and it has been suggested that the term applied to a junior grade of *nautai* or regular seamen.

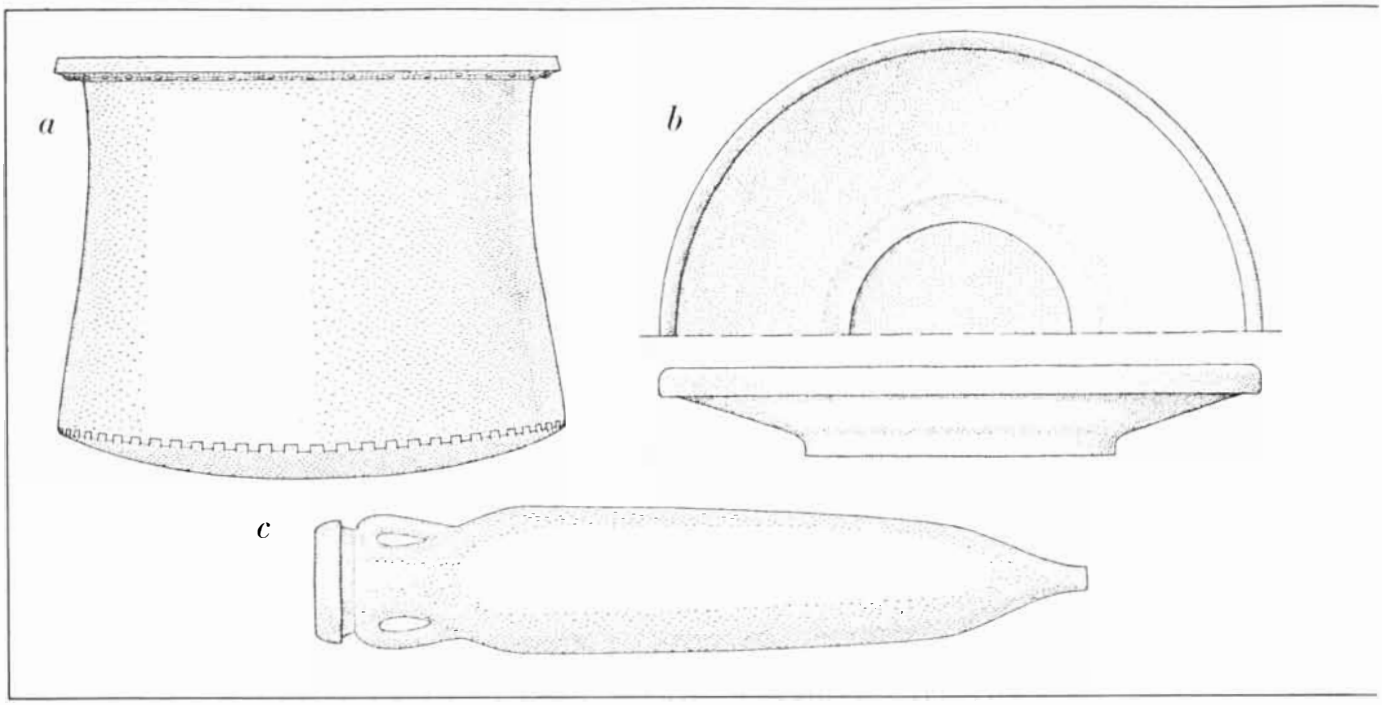
Our ship had steering oars and therefore surely carried a helmsman, but no evidence of his presence survived in the wreck. The same is true of the post of prow officer, unless he was in charge of the anchors (and a conical sounding lead, with a hollow for the tallow or wax that picked up samples of the bottom). The presence of the ship's carpenter is more easily demonstrated. This officer evidently stored his tool chest on deck. A little forward of the galley area at the ship's stern we found a number of shapeless concretions. Here again careful cutting enabled us to cast what proved to be perfect replicas of the tools in a Byzantine tool chest. Michael L. Katzev, who made casts of the hollow cores and analyzed the collection, found that it included an axe, adzes, a hammer-adze, a claw hammer, hammers for metalworking, chisels, gouges, punches, files, drill bits, dividers, an awl, assorted knives and numerous nails and tacks [*see bottom illustration on next two pages*]. A folded sheet of lead and some waste from lead casting that we also recovered from the wreck suggest that various lead fittings could be made on board ship; this may have been part of the carpenter's duties.

If Byzantine boatswains were responsible for the ship's boat, then the grapnel we found is evidence of a kind that Georgios' crew included a boatswain. As for any ordinary seamen who may have served aboard, no evidence of them has survived. In American days of sail, however, square-rigged ships only slightly smaller than Georgios' merchantman often had a crew of no more than three men and a boy. There need not have been many ordinary seamen in the crew to eat further into the profits from the venture. Indeed, there may have been none.

ISLAND OF YASSI ADA, where the ship went down, lies off the southwestern coast of Turkey. It is between the mainland and the larger island of Kalymnos, which is some 20 miles to the west of the ancient city of Halicarnassus. The vessel was evidently sailing from north to south, perhaps bound for one of the celebrated wine-producing centers in the area.

Whether or not *parascharites* stands for cook, there was certainly a cook aboard the ship, and the galley where he worked contained some of our richest finds. Located at the extreme stern of the ship, it was set as low as possible within the hull and was separated from the cargo hold by a bulkhead scarcely eight feet forward of the sternpost. The galley roof stood two and a half feet above the level of the deck and was made of tiles, including one tile with a circular hole that was probably to let the smoke from the galley fire escape. The precise arrangement of the hearth where the fire burned has not yet been reconstructed to everyone's satisfaction. It was built of tiles and iron bars, and it stood in the port half of the galley. Perhaps the tiles formed a low firebox, open at the top and covered by a grill of iron rods that supported the round-bottomed pots used for cooking.

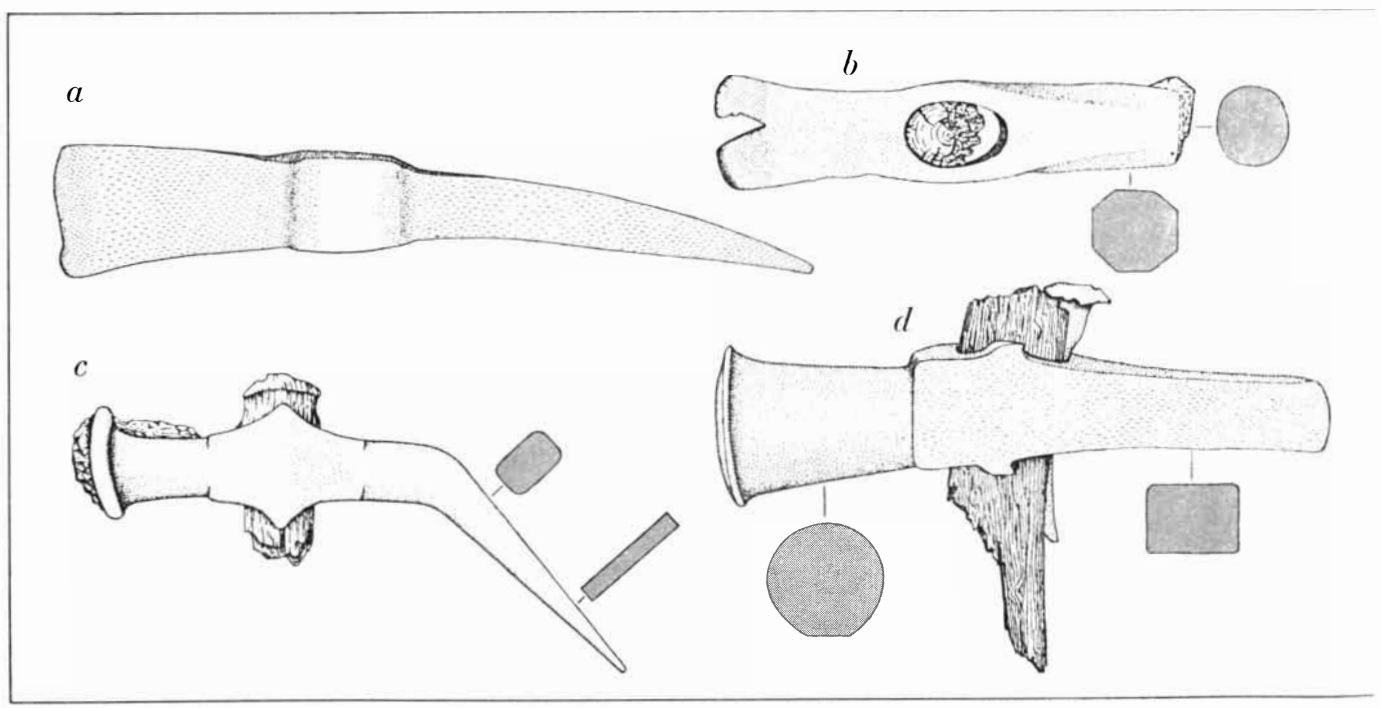
The cook had at his disposal 22 such pots, two copper cauldrons and a large water jar that was stowed at the extreme rear of the galley. A further array of 17 storage jars of various shapes and sizes, doubtless including one or two filled with lamp oil, was kept in a kind of pantry to starboard and forward of the hearth. Elsewhere in the galley, perhaps on shelves and perhaps also in a cupboard of which only a bronze handle has survived, were eight large red plates, two cups, three jars with spouts (two with lids) and 18 pitchers. Most of the pitchers were coated inside with resin, which indicates that they were used to hold wine. Georgios was evident-



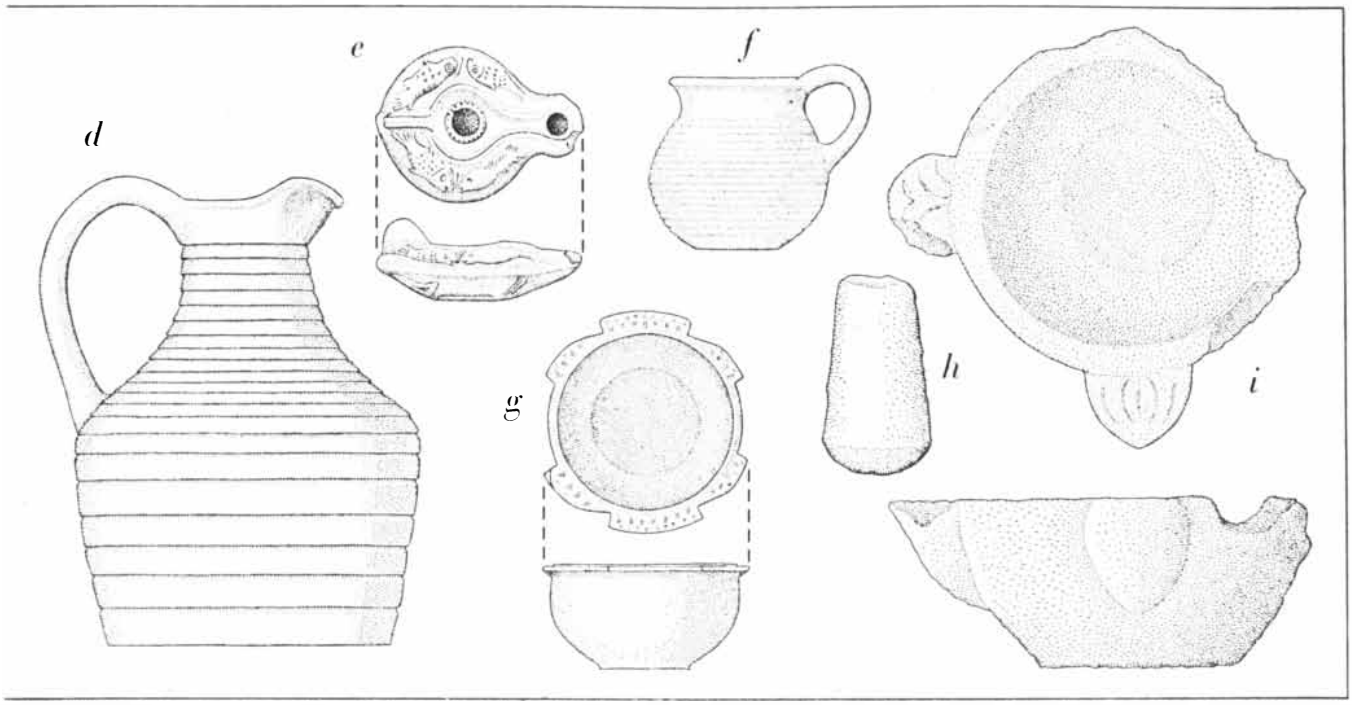
GALLEY EQUIPMENT included several copper cooking utensils. One, a cauldron (*a*), had a bottom ingeniously dovetailed to the body. Eight of the large red-ware dinner plates (*b*) were found in the wreck, as were three of the slender amphoras (*c*), possibly containers for cooking oil. A jug (*d*) that was coated with resin to seal the porous clay was probably used to hold wine. An oil lamp typical

ly not laggard in the fashions of the day: the oldest precisely dated Byzantine lead-glazed pottery consists of four small bowls that we found in the galley area.

The galley may have doubled as a storage area for the officers' gear or indeed as a kind of wardroom. It was in this part of the wreck, clearly separated from the carpenter's kit, that we found the grapnel and a group of iron tools: axes, a pickaxe, pruning hooks and a shovel. These look like boatswain's stores that would have been needed by a landing party to collect firewood for the cook and to enlarge springs to obtain fresh water. The galley area also contained, probably originally stored in the cupboard, a supply of money, a set of balance-pan weights, a bronze censer with wick pin, three bronze belt buckles and a metal belt-tip sheath, the bowl of



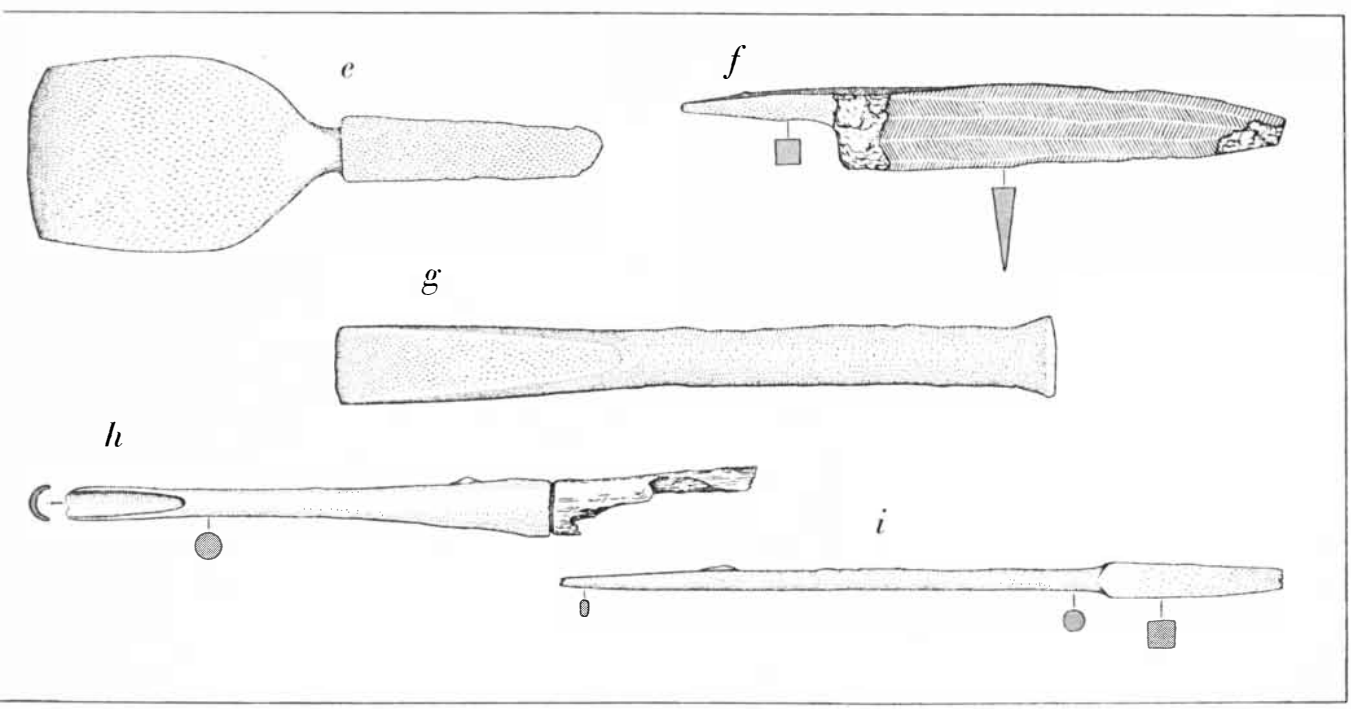
CARPENTER'S TOOLS, destroyed by corrosion, were replicated by making rubber casts of the cavities within the concretions that had formed around them. Among other tools, the ship's carpenter had brought with him a mattock (*a*), a claw hammer (*b*), a ham-



of the 24 found in the wreck (*e*) had been decorated with fish. The single-handled drinking cup (*f*) is one of two such cups that were recovered. The bowl (*g*) is one of four made of glazed ware found aboard the ship and is the earliest precisely dated example of lead-glaze ware from Byzantine times. A stone pestle (*h*) and a matching stone mortar (*i*) were also among the furnishings of the galley.

a lead spoon, a copper tray and several copper vessels. Other cook's gear included a whetstone, a milling stone and a stone mortar and pestle.

What can be said about the ship's last voyage? We have noted that Georgios, in addition to being the captain, may have been the owner and a merchant-venturer as well. The basis for this conjecture is the fact that the scale bearing his name is an item of merchant's equipment. It should be mentioned that later Venetian statutes required that shipowners supply every vessel with a weighing device; perhaps even during Byzantine times a captain may have needed to carry his own balance to show that his freight charges were correct. If we may assume, however, that Georgios was owner as well as master, we can calculate with some confidence his total in-



mer-adze (*c*) and a metalworking hammer that fitted its handle loosely and had been fixed in place with a nail (*d*). Other tools included a hand adze (*e*), a file (*f*), a chisel (*g*), a gouge (*h*) and, to judge by its haft, a wood-boring bit for a brace (*i*). All are half-size.

vestment in the ship and all its stores.

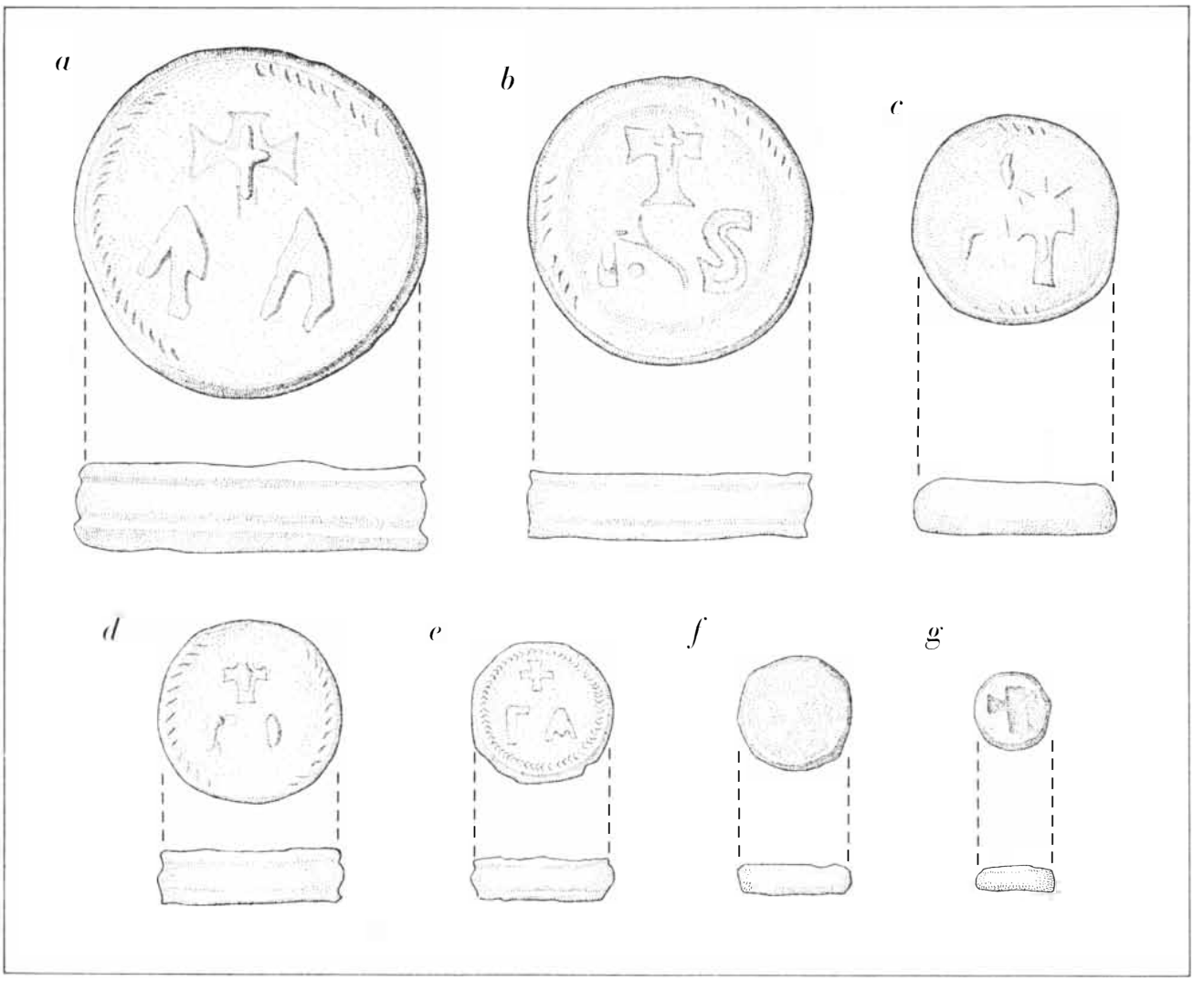
The Rhodian sea law indicates that the seventh-century cost of fully outfitted shipping ran about 50 *solidi* (gold coins) per six and a half tons' capacity. On this basis Georgios' investment would have been some 300 *solidi,* a substantial sum in times when a shipyard caulker might earn 18 *solidi* for a year's work and less skilled laborers might receive only seven or eight. Might this investment have been all Georgios could afford, leaving him with no capital to invest in a cargo? A shipload of wine, for example, would have cost 200 or 300 *solidi* more. One piece of evidence suggests that such was the case.

We found a single lead seal in the wreck; it bears a cross-shaped monogram that we read as the name Ioannes, or John. Such a seal could of course have belonged to a passenger or crew member, but the fact that it was the only seal we found suggests at least the possibility of more official use. Perhaps Ioannes was the *emporos,* or merchant, aboard Georgios' ship, traveling with him to pay the freight charge and handle the sale of goods at the vessel's destination.

If Ioannes was indeed a merchant or merchant's agent, he would have asked other merchants who had sailed previously with Georgios whether or not, as specified by law, the ship was in good condition. When he had heard that it was, he and Georgios would have entered into a contract. Just where Ioannes or his employer had raised the money to buy a cargo was no one else's affair as long as he could pay Georgios the required freight charge.

We know exactly what external shape the cargo took. Loaded aboard the ship were 900-odd amphoras, or storage jars, most of them large and globular but some smaller and more elongated. The large jars could hold as much as 40 liters of liquid, the small ones around nine liters. We cannot be sure what, if anything, the jars contained. Not all of them were lined with resin, the customary method of waterproofing the porous clay. However, if all the jars had been filled with liquid, say wine, the cargo would have weighed some 37 tons. In any event, we can visualize a procession of porters carrying aboard the amphoras, full or empty, and passing them down through the hatch into the hold.

The cook's stores would have been loaded at the same time. Presumably he saw to it that his large jar had been filled with fresh water and checked his supply of lamp oil. We know that his fresh rations included a basketful of



WEIGHTS for a balance scale are made of bronze inlaid with silver. They range from one pound (*a*) to a seventh of an ounce (*g*). The ounce was 20.45 grams; a pound consisted of 14 ounces. They were not standard weights, and their purpose is not known.

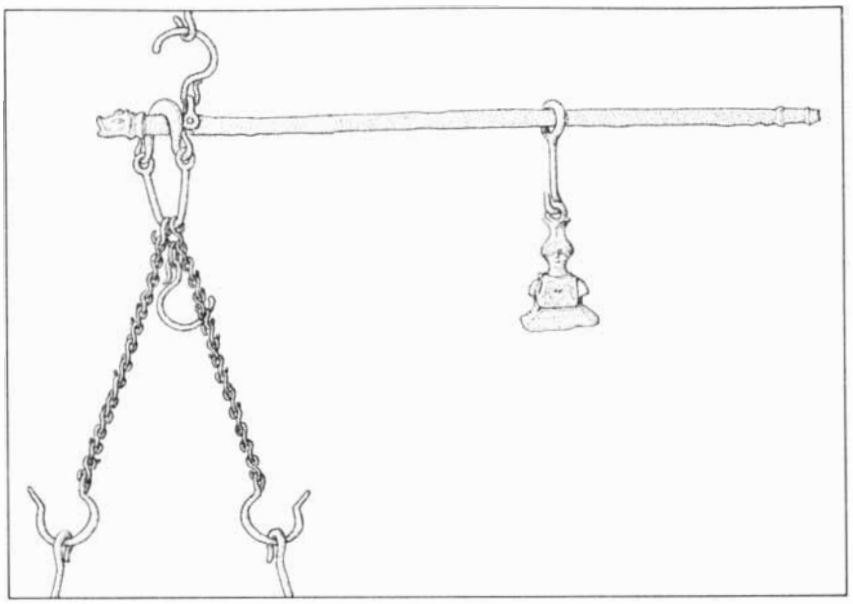
dark, gleaming Bosphorus mussels: we found piles of their shells in the wreck and they are not native to these waters. We know too that the bow officer—if such was his responsibility—had lashed a pair of anchors to the port gunwale and another to the starboard gunwale near the bow and had stowed the remaining seven just forward of the mast.

Meanwhile the captain placed in the galley certain valuables, including a money purse or two. We found 54 copper *folles* (coins worth a small fraction of a *solidus*) and 16 small gold pieces; the total value of the coins was just a little more than seven *solidi*. Was this a ship's fund or money deposited by a passenger? The Rhodian sea law declares: "If a passenger comes on board and has gold, let him deposit it with the captain. If he does not deposit it and says 'I have lost gold or silver,' no effect is to be given to what he says, since he did not deposit it with the captain." If one is to try to answer the question, one needs to know something about the purchasing power of the *solidus*.

As we have noted, seven *solidi* represented a year's wage for some kinds of labor; early in the eighth century, for example, a blacksmith earned three-quarters of a *solidus* per month. Early in the seventh century a cloak might cost from one to three *solidi*, and later in the century one *solidus* would buy four cheap blankets. Food was much less dear. One sixth-century figure gives five *solidi* as the cost of a year's rations for one man; in the years when our ship made its final voyage a loaf of bread cost three *folles*. The money found in the wreck would have fed a crew of 15 for a month with something left over, not even taking into consideration the evidence (lead weights for fishing lines and nets) that the ship's company supplemented its provisions en route. It therefore seems likely that the contents of the purse or purses were the ship's victualing money.

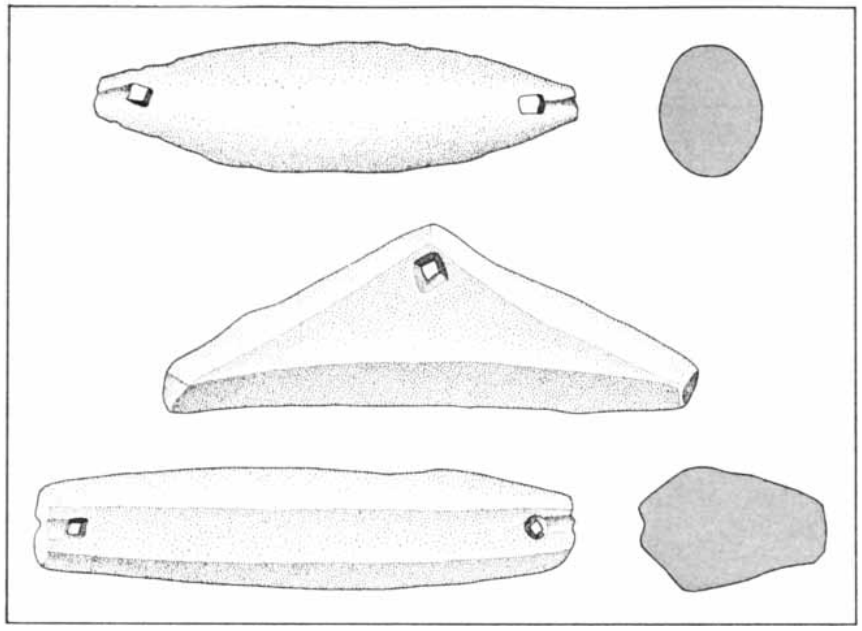
The 70 coins, by the way, are what enable us to pinpoint the date of the voyage at A.D. 625 or 626. Six of the coins were too badly preserved to allow identification. Of the remainder Joan Fagerlie of the American Numismatic Society finds that only two were minted earlier than the reign of the Emperor Heraclius (A.D. 610–641). The latest coin in the group was minted in the 16th year of the emperor's reign, that is, in A.D. 625. We may safely assume that the ship sank in the same year or quite soon thereafter.

The weighing equipment for the voy-



CAPTAIN'S SCALE made of bronze was of the fulcrum-and-counterpoise type later called a steelyard beam. The object to be weighed was suspended from the short end of the beam, and the counterpoise was moved outward until the scale balanced. The weight was then read from calibrations along the beam. Scale was calibrated for both heavy and light loads.

COUNTERPOISE for the steelyard beam was a hollow bronze bust of the goddess Athena weighted with lead. The Gorgon's head on the breastplate identifies the goddess. Athena's image was often used in this way even though the Byzantine Empire was officially Christian.



FISHING WEIGHTS found in the wreck were used with both lines and nets. Their presence implies that the crew members fished during the voyage to supplement their rations.

age was also stowed in the galley cupboard or close to it. The balance-pan weights, made of bronze inlaid with silver, came in a wooden tray that held them in a graduated series in cylindrical pockets. Seven of the original nine weights were recovered. In addition we found what may be a fragment of the balance itself. The bronze steelyard beam that bears Georgios' name was one of three scales of this kind in the wreck; one of the other two was made of iron and was therefore poorly preserved. All this equipment has been studied in detail by G. Kenneth Sams of the University of North Carolina.

Georgios' steelyard beam is decorated with a boar's head at one end and the head of another animal, possibly a dog or a lion, at the other [*see top illustration on preceding page*]. It has two fulcrum points and two calibrated scales, for heavy or light loads; the counterweight is a lead-filled bronze bust of the goddess Athena. (Two similar counterweights in Athena's image have been discovered in wrecks, one off Sicily and the other in the Bosphorus.)

Both bronze steelyard beams are calibrated in terms of a pound that was equal to 315 grams. (Our pound troy, which is 373 grams, is the closest modern equivalent; our pound avoirdupois is 453 grams.) Van Doorninck has found that the same lightweight pound was the unit of measurement for the ship's anchors. The same unit is very close to one determined earlier from weights of fifth-century and sixth-century Byzantine times.

Curiously the balance-pan weights represent an entirely different system. Most are clearly marked; they include a one-*litra,* or one-pound, weight; a six-*unciae,* or six-ounce, weight, and three-ounce, two-ounce and one-ounce weights. Two smaller weights are unmarked: one of three *nomismata* (six *nomismata* normally equaled one *uncia*) and the other of one *nomisma.* This is possibly the most complete set of Byzantine weights in existence, and their pound is not the standard "light" 315-gram one. It is superlight: 287 grams. Furthermore, it is a pound divided into 14 ounces rather than the customary 12, and its ounce is divided into seven *nomismata* rather than six.

These unusual values are not the result of any alteration in the original weights as a consequence of their long submersion. Each of the weights in the set, give or take a fraction of a gram, is consistent with all the others; in every case the ounce is 20.45 grams. It happens that an ounce of this weight was the Byzantine standard for gold coinage; the standard was established in the time of Constantine the Great (A.D. 272–337) on the basis of the old Roman pound (327.45 grams). In this coinage-weight system, however, there were 16 ounces to the pound rather than 12 or 14. Only one other 14-ounce pound is known: during the fourth century a "heavy" pound divided into 14 ounces rather than 12 was used in the mining of gold in order to increase state revenue from mine leases. All of this, however, can have nothing to do with the standard of the balance-pan weights from the wreck. At present it remains unexplained.

The wreck contained one more balance weight. It is a small pendant of yellow glass that is pierced for stringing. It is so similar in appearance to other Byzantine glass weights that it must have been one. The pendant is shaped like a coin, and pressed into the glass is another cruciform monogram; this one gives the name Theodoros. Such weights are thought to have been used to check gold coins, although the weights of the glass pieces seldom correspond to the weights of any coins then in circulation. The discrepancy is explained on the grounds that the glass pieces have lost weight as a result of corrosion.

So far we have seen Georgios' ship built, supplied, loaded and under way. What was its home port, its course and its intended destination? It seems safe to state, on the basis of the coins alone, that the voyage was from north to south. With a single exception every one of the identifiable copper coins had been struck at mints in the northern part of the Byzantine Empire: Constantinople, Thessalonika, Nikomedia, Cyzicus. Only one copper coin, and it is one of the oldest, is from a mint in another part of the empire; it was struck at Alexandria ad Issum on what is now the Levantine coast of Turkey. Since the copper coins are overwhelmingly northern, the home port of the ship should probably be sought in the same direction.

The pottery from the galley strengthens this assumption. Karen D. Vitelli of the University of Maryland at Baltimore has found that fully half of the oil lamps we recovered are "Asia Minor" types. Seven are of a kind common in Bulgaria, Romania, Thrace and the Hellespont; others find their closest counterparts at Samos, Ephesus, Smyrna, Troy, Miletos, Delos and Chios. A probable origin cannot yet be assigned to the rest of the lamps, but as with the coins few if any appear to have come from the parts of the empire west, south or east of the shipwreck. Other wares from the galley also find their best parallels to the north. The lead-glazed bowls speak of Constantinople, and other pottery is similar to the ware of Chios and of two Black Sea ports: Histria and Tomis.

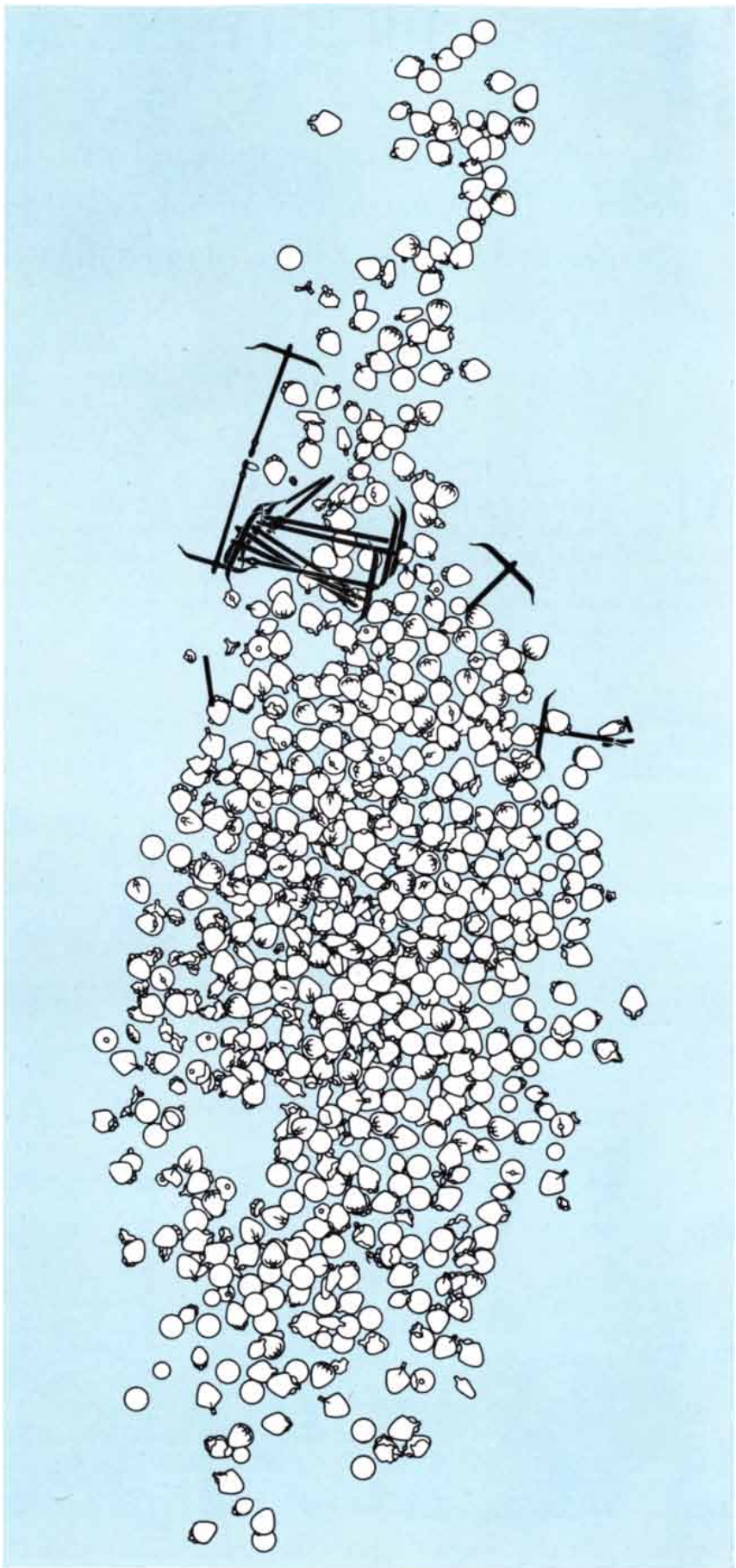
The presence of Bosphorus mussels among the ship's victuals provides still another piece of evidence suggesting a

northern home port for the ship, perhaps in the Hellespont itself. Finally, the place where the ship went down is compatible with a course running from north to south. The wreck lay on the southeast side of a treacherous reef off a small coastal island named Yassi Ada. Throckmorton, who found the ship, is of the opinion that it was running southward before a wind from the north when the reef tore a hole in its bottom; evidently it sank so fast that there was no time to salvage valuables from the galley or to drop even a single anchor.

In the light of all this evidence it seems reasonable to assign the ship a home port not too far from Constantinople. It is virtually certain that the vessel was on a southward course when it sank. There remains the question of its destination. Today one sees vessels plying the Aegean whose hold and deck are crowded with pottery containers. The pots are empty; they are the cargo.

We have seen that many but not all of the amphoras aboard Georgios' ship were lined with resin. A potful of congealed resin was also found in the galley. This suggests that resin was being melted during the voyage; perhaps the purpose was to coat the unlined amphoras. Moreover, although the ship carried some 900 amphoras, we found in the wreck only 100 amphora stoppers. (It is of course possible that the missing stoppers were made of some perishable material.) The ship's course was carrying it toward Kos, Knidos and Rhodes, all of them celebrated wine-producing centers. It is not illogical to suggest that Georgios was carrying a load of the best wine jars available on the Constantinople (or Black Sea) pottery market for the use of vintners in the south, and that the merchant Ioannes had decided to cut costs a little by having some of the unlined pots coated with resin on the voyage.

There are of course other questions, some of which may be answered and some not. The main thing we have learned is how much a single shipwreck—and not a particularly well-preserved one—can tell us.



CARGO OF AMPHORAS littered the sea floor off Yassi Ada as they spilled from the decaying shipwreck. More than 900 were recovered by the excavators. The majority were globular, with a maximum diameter of about 18 inches. A few, more elongated in shape, were only 10 inches in diameter. If all had been filled with liquid, the weight of the cargo would have been some 37 tons.

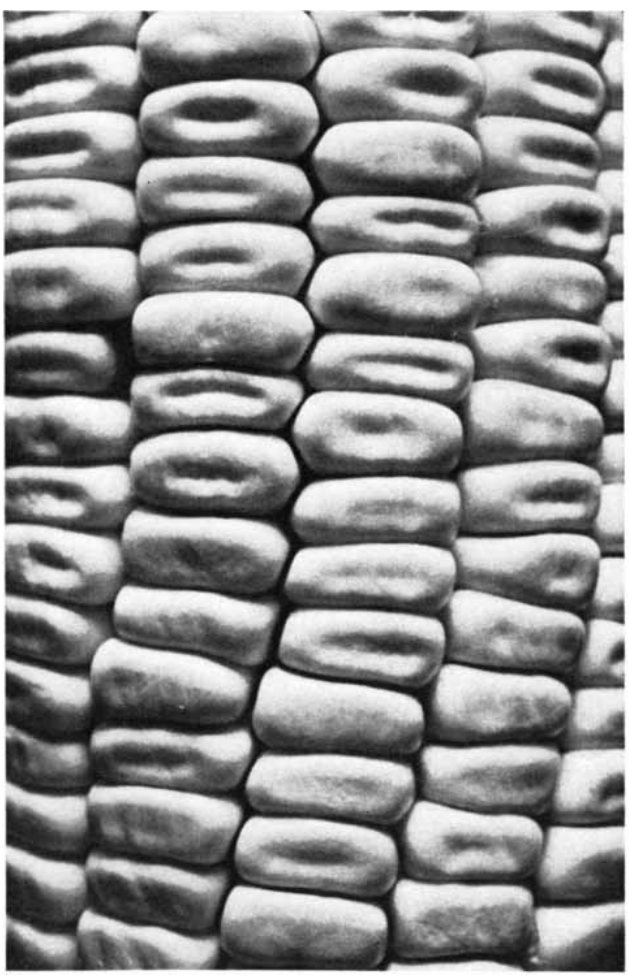
HIGH-LYSINE CORN

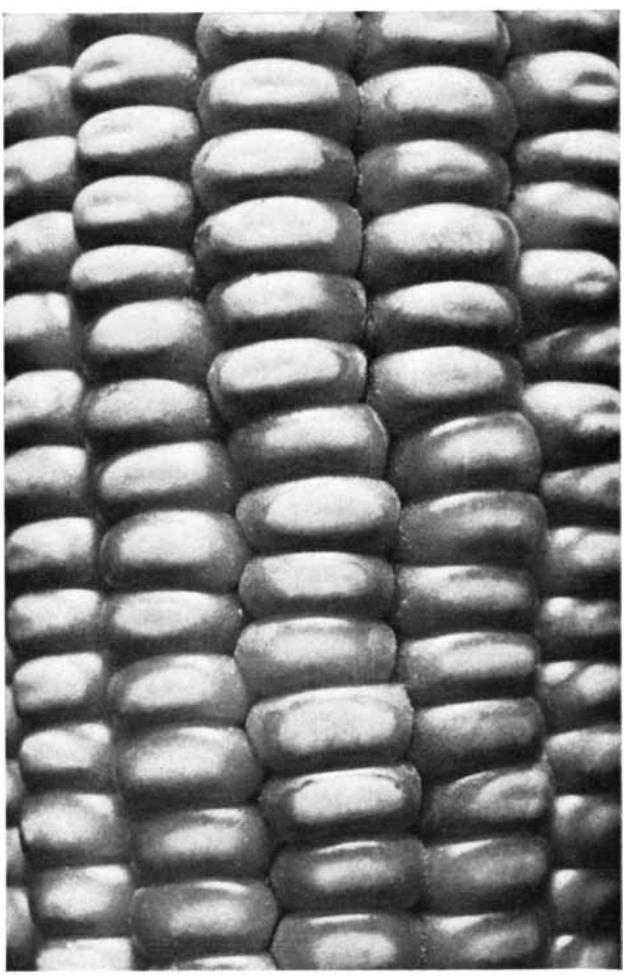
As a source of protein for men and other nonruminant animals corn is deficient in the amino acid lysine. This deficiency is now being remedied by the breeding of high-lysine strains

by Dale D. Harpstead

Mankind depends on grain crops for 70 percent of its consumption of protein, the most essential food. For the vast populations of the developing countries, unable to afford the luxury of animal protein, cereal grain is practically the only important source of protein in the daily diet. One of these dietary mainstays is corn; in total world production it is outranked only by rice and wheat, and it is the principal staple in many countries, particularly in the Tropics and sub-Tropics. Advanced countries, notably the U.S., also rely heavily on corn, as it is the main feed used for the production of animal protein in the form of meat, milk and eggs.

Yet corn is a poor source of protein for a single-stomach animal (including animals such as the pig and the chicken as well as man). The protein of corn is not





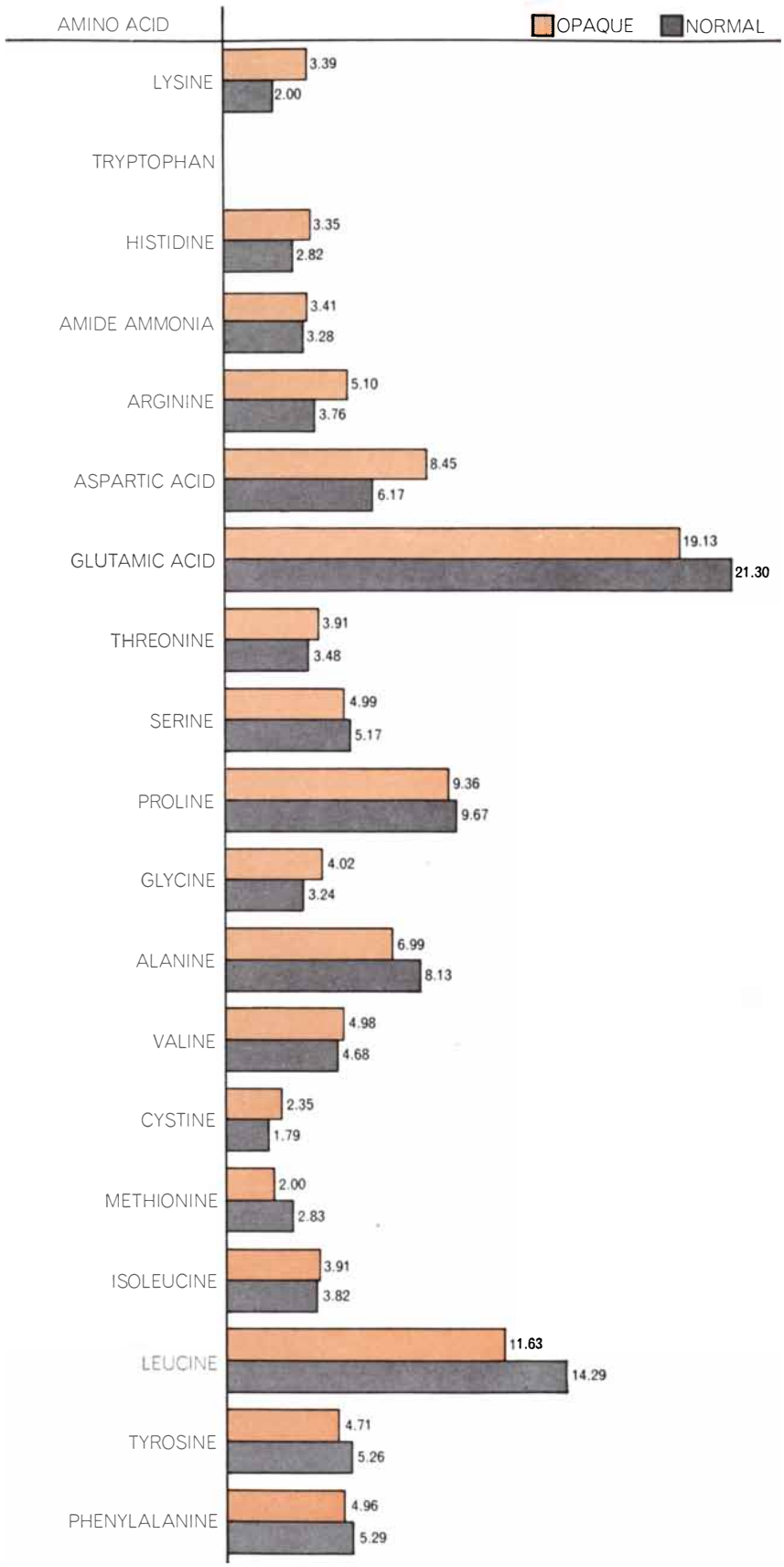
KERNELS OF HIGH-LYSINE CORN (*left*) are compared with those of a normal tropical flint corn (*right*). The endosperm within the kernels of the high-lysine ear is opaque and floury, whereas the endosperm in kernels of the normal ear is translucent and darker.

only low in quantity (about 10 percent); it is also low in quality as food. About 50 percent of the protein in corn is zein, which cannot be digested by a nonruminant (single-stomach) animal. Corn's most serious deficiency in terms of nutritive value is its low content of lysine, an essential amino acid that man and other nonruminant animals do not synthesize and must obtain from their food.

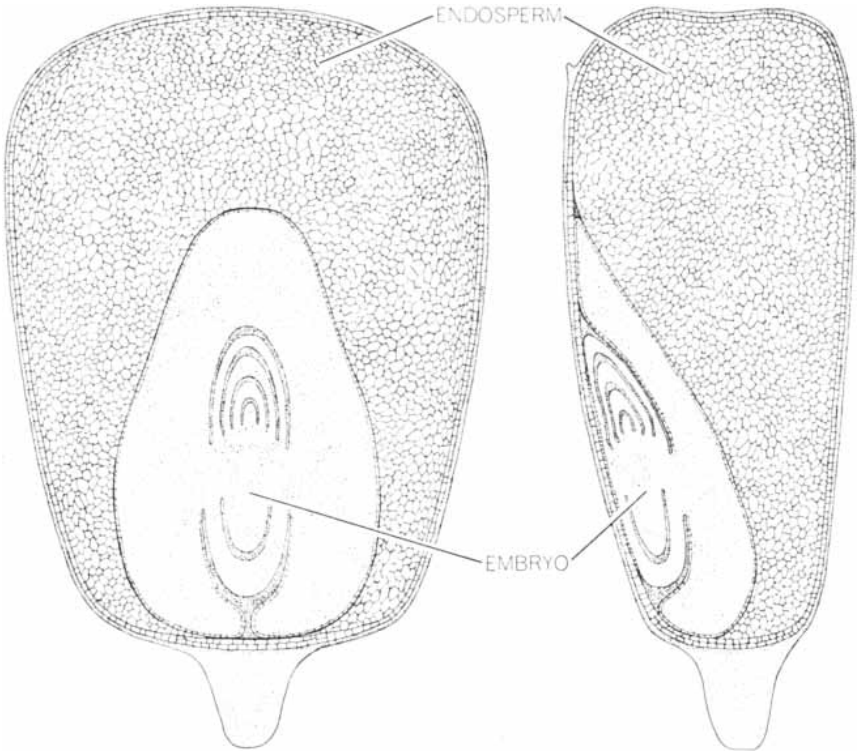
Many years ago Thomas Burr Osborne and his associates at the Connecticut Agricultural Experiment Station found in experiments with rats that a diet containing corn as the sole source of protein produced obvious signs of malnutrition. They were able to correct the deficiency by supplementing the corn diet with small amounts of lysine and tryptophan, another amino acid that is present in corn only in meager amounts. At the human level the inadequacy of corn as a protein source is shown strikingly by children in countries where corn is the principal food; among these children there exists a high frequency of the protein-deficiency syndrome called kwashiorkor.

Over the decades since Osborne's discoveries plant geneticists have been working to breed varieties of corn with higher protein quality. It has been widely recognized that success in this effort would be a major breakthrough in the worldwide food problem [see "A World Agricultural Plan," by Addeke H. Boerma; SCIENTIFIC AMERICAN, August, 1970]. Up to a few years ago all the breeding efforts met with failure. A high-lysine corn has at last been produced, however, and the breakthrough now seems well under way.

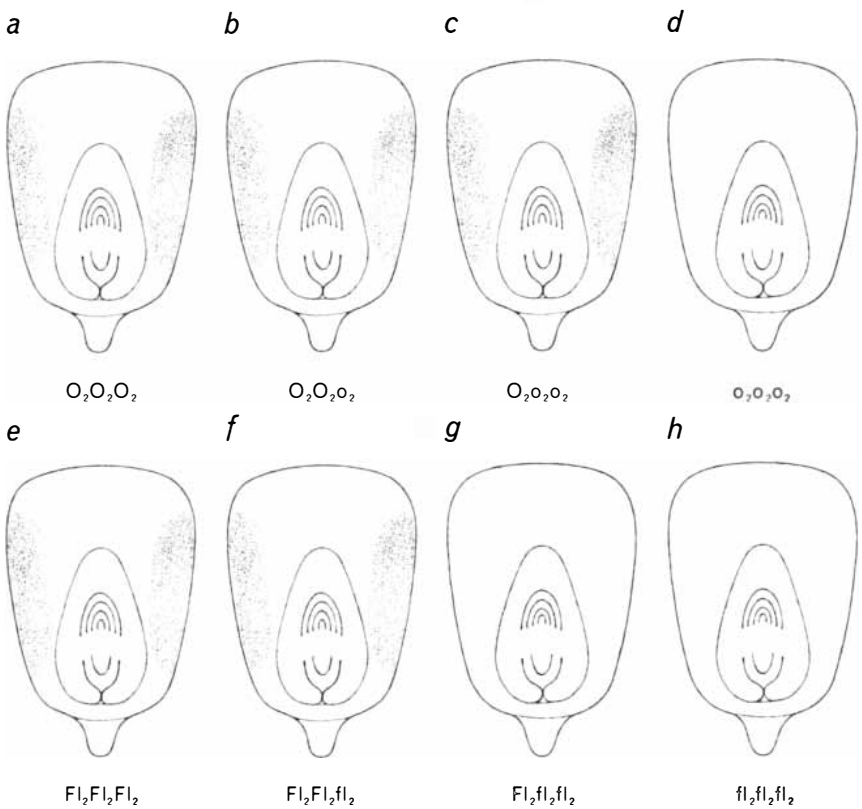
In the 1950's Edwin T. Mertz and Ricardo Bressani of Purdue University began to search for corn variants that might be exceptionally rich in protein. Working with varieties of corn from Central America, they found no significant leads. Oliver E. Nelson, who had been studying starchy mutants of corn, joined the team as a second attempt at protein modification was being planned; the group then turned to certain soft-kernel mutants known as the "opaque" and the "floury" types. The "opaque" mutants were named for the fact that, in contrast to the translucence of kernels in normal corn, the endosperm of these mutants does not transmit light; the "floury" mutants were named for their soft, fine-textured kernels (a property that is also present in the opaque type). The properties are governed by the presence of a recessive gene or genes. In both of these groups of mutants when the homozygous



AMINO ACID CONTENT of endosperm from opaque and normal kernels is tabulated. The units are grams of amino acid per 100 grams of protein. No figures are given for tryptophan because it is destroyed by the hydrolyzing agents used to break protein down into its amino acid units. The measurements were made by Edwin T. Mertz and his colleagues.



ENDOSPERM OF HIGH-LYSINE KERNEL is shown in longitudinal section (*left*) and transverse section (*right*). The endosperm of the normal kernel is dense at the sides.



GENES FOR HIGH LYSINE CONTENT are the recessive "opaque" gene (o) and the recessive "floury" gene (fl). The dominant genes at the same loci on the chromosome (O and Fl) are associated with normal endosperm. It takes three doses of the recessive opaque gene to give rise to a high-lysine kernel (*d*) but only two doses of the recessive floury gene (*g and h*). Three genes are involved because endosperm of corn results from what is called double fertilization, so that cell nuclei contain three sets of chromosomes instead of usual two.

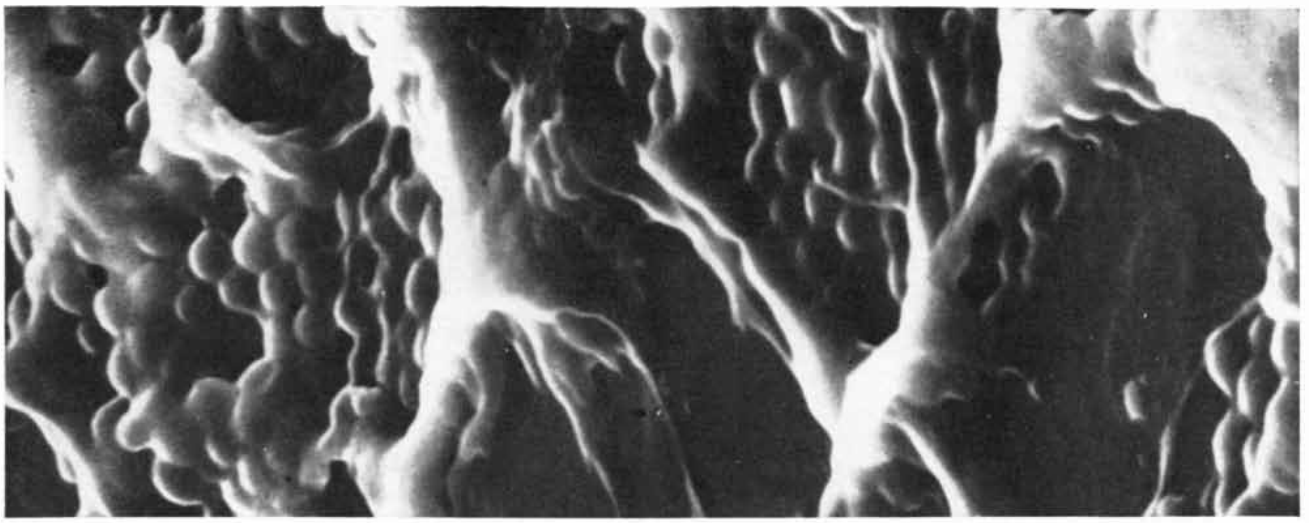
recessive situation prevails (only the recessive form of the gene being present), the endosperm lacks the hard, vitreous body characteristic of ordinary corn, and the kernels look dull and faded in color. These distinctive properties are highly convenient as markers for tracing the gene in breeding experiments.

The Purdue investigators soon turned up a rich find in a gene called "opaque-2" (because it was the second mutant to be discovered in the opaque group, which had been identified and catalogued many years before). Mertz's team found that this gene, when introduced into an ordinary strain of corn, could increase the kernels' lysine content by as much as 69 percent. Shortly afterward they discovered that a gene in the floury series, called "floury-2," also had a similar capability.

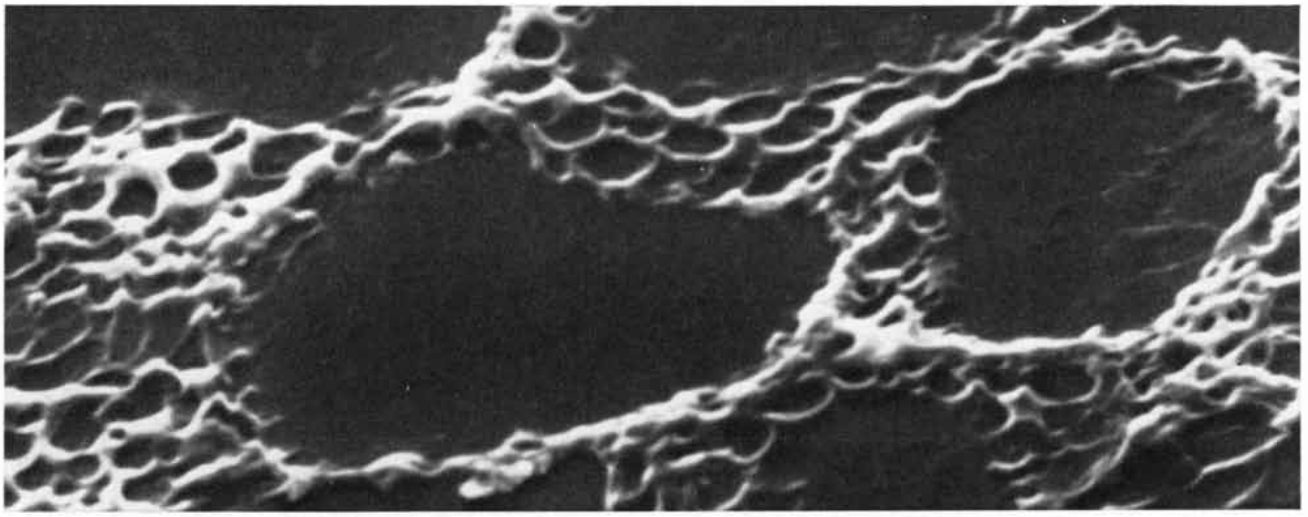
The nutritional proteins in corn kernels are classified in four categories according to solubility: (1) albumins, soluble in water; (2) globulins, soluble in saline solutions; (3) prolamines, soluble in moderately strong alcohol, and (4) glutelins, soluble in dilute alkaline solutions. Of these four fractions the dominant ones in "normal" corn are prolamine, which includes zein, accounting for 40 to 55 percent of the protein total, and glutelin, which accounts for about 30 to 35 percent of the protein total and carries all or almost all the endosperm's lysine. By biochemical analysis the Purdue investigators and others established that in kernels modified by the opaque-2 or floury-2 gene the zein fraction was substantially reduced and the glutelin fraction was correspondingly increased. Thus the lysine content of the endosperm was raised from the ratio in ordinary corn (two grams per 100 grams of protein) to a ratio of 3.39 grams per 100 of protein [*see illustration on preceding page*]. There were also noteworthy increases in tryptophan and several other amino acids.

The Mertz group's discovery of a means of producing high-lysine corn, published in 1964, came as a breath of fresh air to plant breeders and nutritionists. It stirred the imagination of many people who were concerned with the worldwide problem of protein need. The possibility that corn might become an agent for relief of the world hunger for protein appeared to be as important a find as the historic discovery that pellagra could be prevented by niacin.

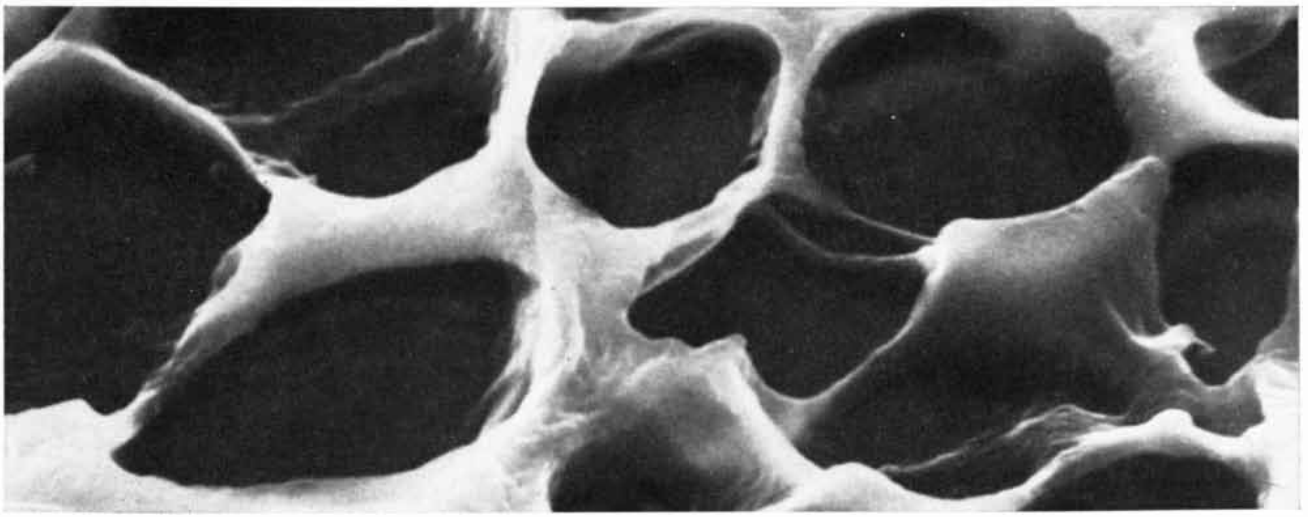
The new corn was soon put to the test of its actual value in nutrition. Preliminary tests with laboratory rats and pigs, made as soon as supplies of the modified corn seed became available, showed that



ENDOSPERM OF NORMAL CORN is enlarged 2,750 diameters in scanning electron micrograph. Starch has been dissolved with enzyme amylase, leaving two protein components: zein (*round bodies*) and matrix proteins. Zein is deficient in lysine and tryptophan.



ZEIN IS REMOVED with hot ethyl alcohol, leaving matrix protein with holes where the zein was. The matrix consists largely of the protein glutelin. The micrographs on this page were made by Uheng Khoo and F. L. Baker of the U.S. Department of Agriculture.



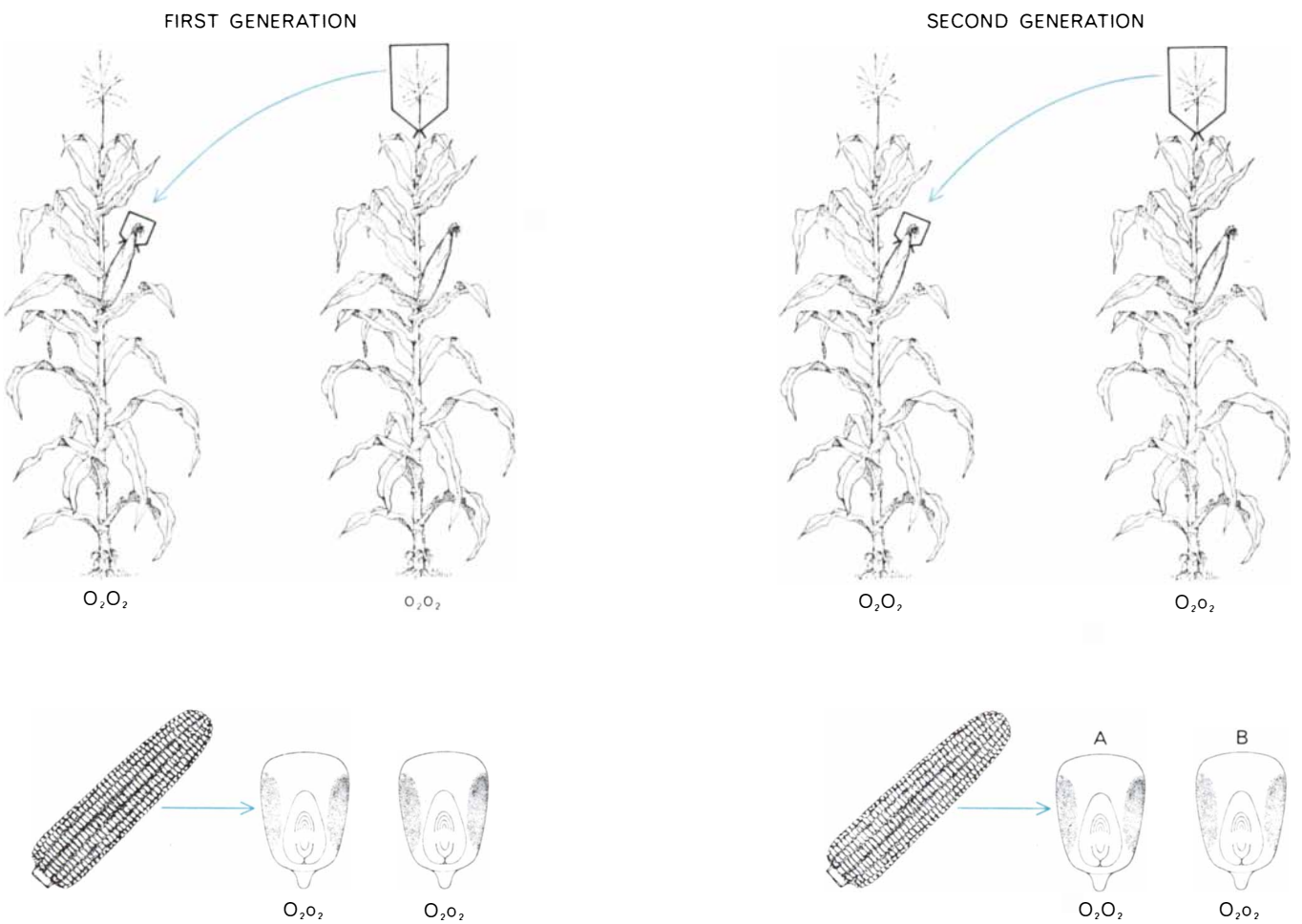
ENDOSPERM OF HIGH-LYSINE STRAIN "opaque-2" is shown with starch dissolved away. Zein component is absent, so that principal protein is glutelin. Since glutelin has a higher proportion of essential amino acids than zein, so does endosperm as a whole.

the high-lysine corn was indeed a dramatic improvement over ordinary corn as a food. It could be assumed that not all the new corn's improvement in nutritive value was attributable to its increase in lysine content. No doubt the increase in tryptophan also made a significant contribution, and so might some of the other amino acids whose ratio was enhanced. At all events, the change of the protein pattern in opaque-2 and floury-2 corn and the nutritive effects demonstrated that the laboratory discovery had in fact yielded a new product quite unlike "normal" corn.

It was my privilege to be working in Colombia as "quarterback" of a team of investigators—plant breeders, animal nutritionists, physicians, grain millers, government officials—that was seeking to improve the protein yield from plants. We had an ideal situation for exploring the potentialities of the new corn. The Colombian government was greatly interested in the protein problem and was willing to support a bold program for solving it. Our team of diverse talents was able to make full use of the available materials and technology for development of the new discovery. And Colombia's tropical climate and year-round growing season allowed rapid progress in the plant-breeding work.

The first requirement was to breed the new genes into normal corn stocks that were suitable for commercial production, so that large-scale tests could be made of the new product's growth potential and value. The introduction of the genes will itself be a demanding enterprise. Corn is not a monolithic plant; it has many varieties, each adapted genetically to its own locality and the use for which it has been bred. Hence in each case the high-lysine gene will have to be introduced into a stock that fits the locality and desired use. For hybrid corn it will be necessary to insert the high-lysine gene in each of the inbred lines that make up the parent stock [see "Hybrid Corn," by Paul C. Mangelsdorf; SCIENTIFIC AMERICAN, August, 1951].

Using as stocks the locally adapted varieties of hybrid corn in Colombia, we proceeded to transfer the opaque-2 or floury-2 gene into the inbred lines by the standard technique of backcrossing. By cross-pollination, selection of the daughter plants found to be carrying the desired gene and repeated crossing back onto the original stock one finally obtains, usually after five or six generations, a seed that carries the new gene and in other respects nearly duplicates the characteristics of the original line. In the case of the high-lysine genes one has the great advantage of detecting the presence of the gene at each passage just by



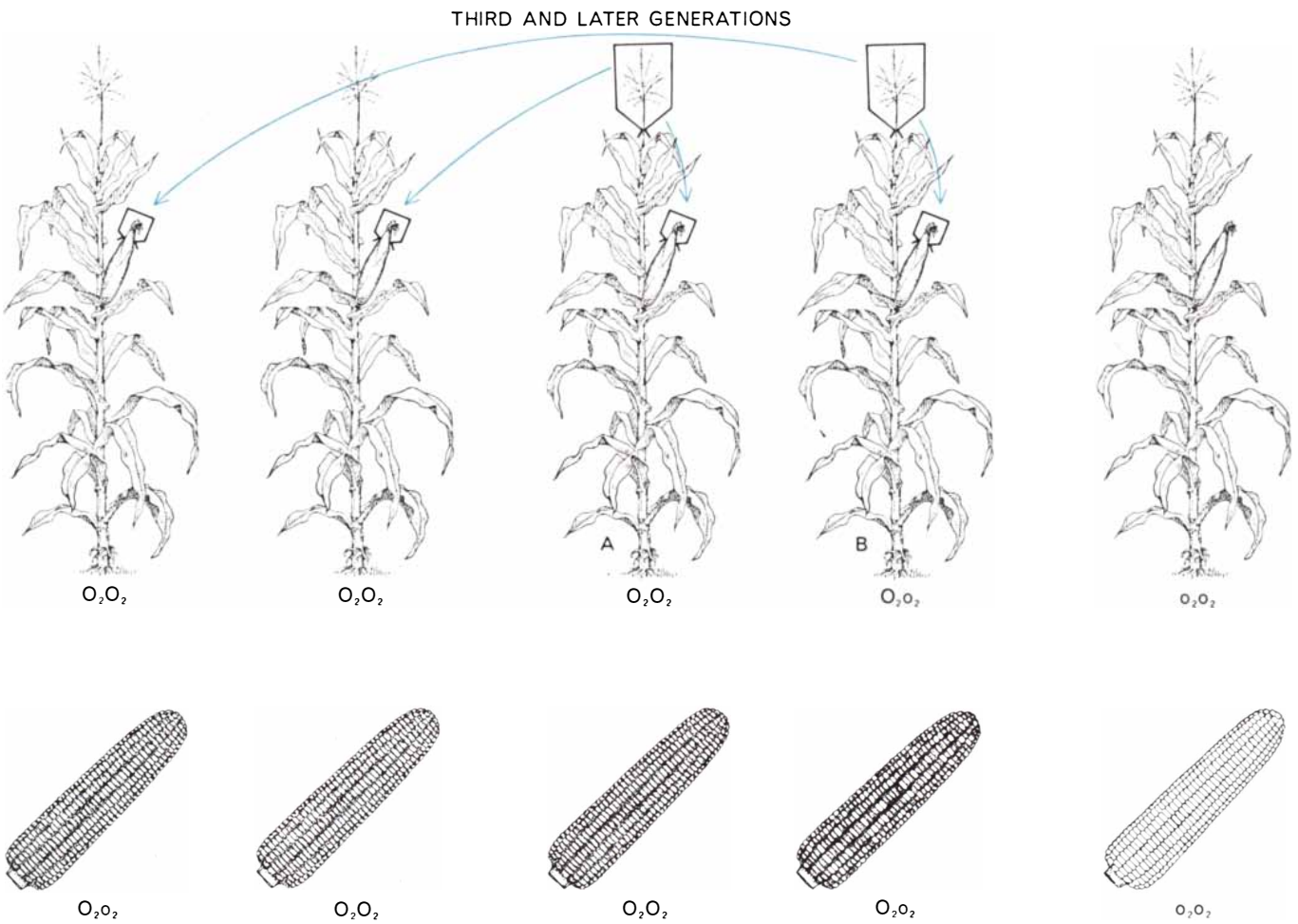
HIGH-LYSINE FACTOR is bred into a locally adapted strain of corn by repeated backcrossing. In the first generation pollen from a high-lysine strain (o_2o_2) is used to fertilize an ear of the locally adapted strain (O_2O_2). All the kernels on the fertilized ear are heterozygous (O_2o_2). In the second generation the seed is used to grow heterozygous plants, pollen from which is again used to fertilize ears of locally adapted plants. The kernels from these ears are half homozygous for the local strain (O_2O_2) and half heterozygous (O_2o_2), but there is no difference in appearance between them. In the third generation the homozygous seed (*A*) and the heterozygous (*B*) are used to grow new plants. When pollen from one of these plants is used to fertilize one of its own ears, the appearance of the

visual examination, since it shows its presence by producing soft, floury kernels. With this marker present, one knows that the kernels carry the improved protein without having to make a chemical analysis or biological assay. A technique involving both cross-pollination and self-pollination of the male parent makes it easy to identify plants that will transmit the recessive gene [*see illustration below*].

By January, 1967, we had enough high-lysine corn to start large-scale trials of its nutritional effects. For tests in animals the pig is a most useful subject: its digestive system is very similar to man's, its feeding requirements are well known, and these requirements are themselves a matter of great interest because the development of a low-cost, high-quality feed for hogs would have considerable economic importance. The complex mixed feeds used for hog-raising are costly.

The effect of a given feed on the pig can be measured rather quickly and precisely, as there are well-established standards. A pig raised on an adequate, balanced diet will reach a weight of 200 pounds (90 kilograms) in 156 days. Its growth is defined in two stages: rapid early body-building, up to a weight of 110 pounds (50 kilograms), and the "finishing" stage of fattening to 200 pounds. Richard A. Pickett of Purdue had already found in preliminary tests of the new corn that piglets grew three and a half times faster on high-lysine corn than on normal corn as the sole protein source. The ratio of units of feed required per unit of weight gain was only 3.3 to one for the high-lysine corn as against seven to one for normal corn. And young pigs did about as well on high-lysine corn as on a mixed feed of normal corn and soybean meal supplying the same amount of nitrogen. The Purdue investigators went on to test high-lysine corn as a finishing feed and got essentially the same comparative results. During this stage, when the pig's growth rate is not quite as rapid, the pigs on high-lysine corn grew 50 percent faster than those on normal corn without a supplement.

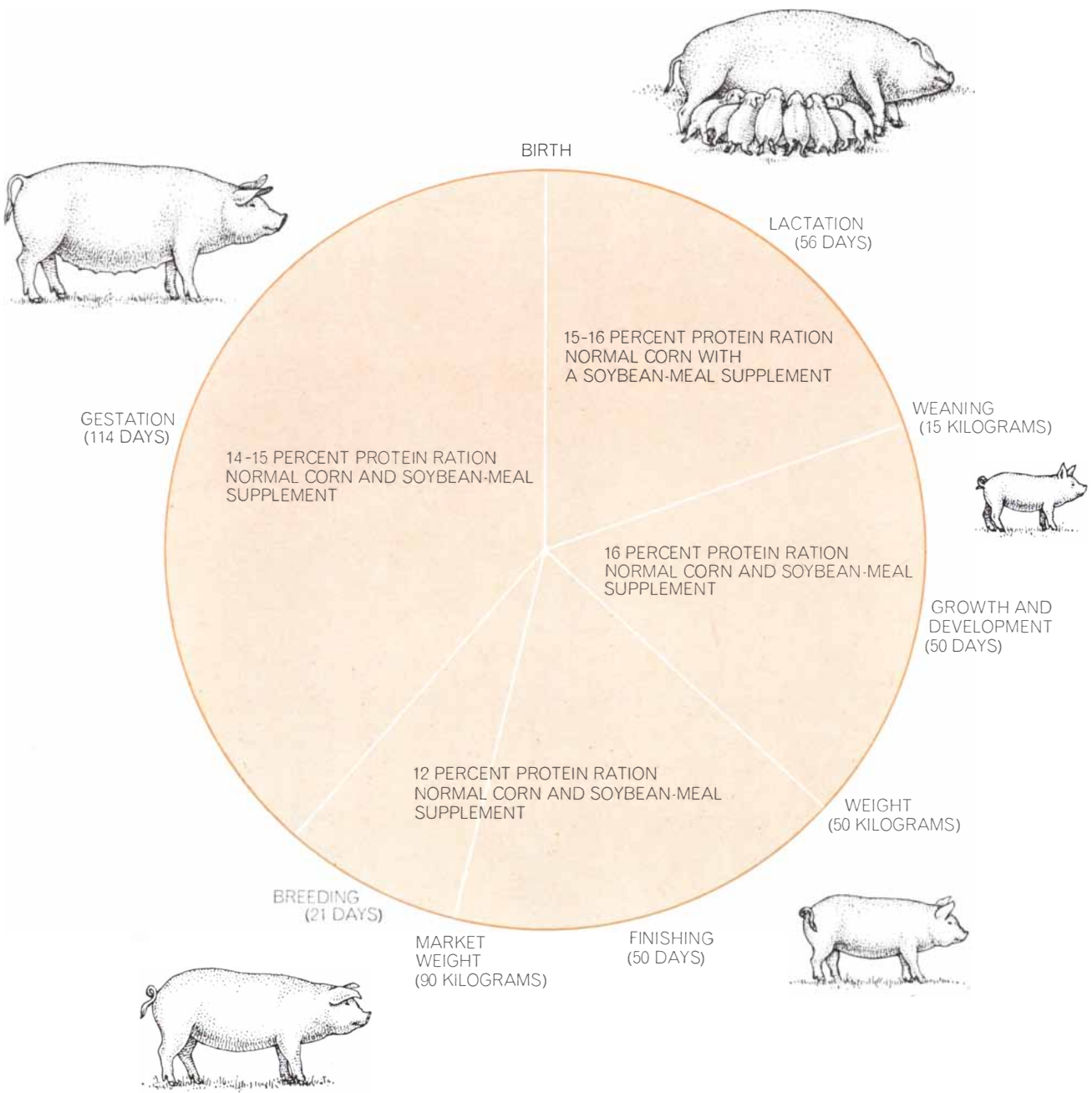
In Colombia, Jerome Maner of our team in collaboration with members of the Colombian animal husbandry institute set out on a systematic study of the comparative effectiveness of high-lysine corn as measured against normal corn with various amounts of protein supplement. They tested the experimental diets with weanling pigs 35 days old and weighing 20 pounds. After 130 days it was found that the pigs on high-lysine corn had gained 73.2 pounds (an average of 256 grams per day) whereas the



ear now indicates whether it is homozygous or heterozygous. The kernels of the homozygous ear all look the same, and the ear is discarded. Moreover, now that it is known that the plant is homozygous, the ears from backcrossing of the same plant and the homozygous locally adapted plants can also be discarded. The kernels on the ear of the self-fertilized heterozygous plant, however, segregate into normal kernels and the opaque high-lysine kernels; the ratio of segregation is three normal kernels to one opaque kernel. After five or six generations of backcrossing, which is necessary to ensure that most of the other genes are those of the locally adapted strain of corn, opaque seed from the segregated ears can be used to grow plants that are homozygous for the high-lysine factor (*far right*).

weight gain of pigs on unsupplemented normal corn was only 6.6 pounds (21 grams per day). Even the high-lysine corn supplies substantially less protein than an optimal diet for pigs. Nevertheless, the pigs fed the new corn were healthy, were normal in general bodily appearance and were inferior to optimally fed animals only in a slight lack of development in some internal tissues. On the other hand, the pigs on the diet of normal corn without supplement showed atrophy of essential organelles within the cells, stunting of the skeletal system and fatty degeneration of the liver tissues. These defects in the pigs were similar to those found in human beings who have suffered severe and prolonged protein deficiency.

The series of tests conducted by Maner and his associates demonstrated that the weight gain achieved with high-lysine corn fell substantially short of that produced by an optimal feed of corn and soybean meal supplying a balance of essential nutrients and protein at the 16 percent level (including nonusable as well as usable protein). There is practical wisdom, however, in a Spanish proverb that says, "The best is an enemy of the better"—meaning that reaching for the unattainable may sometimes thwart improvement. In most of the developing countries to discuss optimal diets for man or animal is in fact utopian, as far as large segments of the population are concerned. In Colombia, for example, the availability of fish meal or soybean meal as a corn supplement is limited and cannot be counted on. Enriched feeds for pigs are beyond the economic reach of the small farmer. Therefore our research has focused on the objective of using high-lysine corn to obtain a high yield of

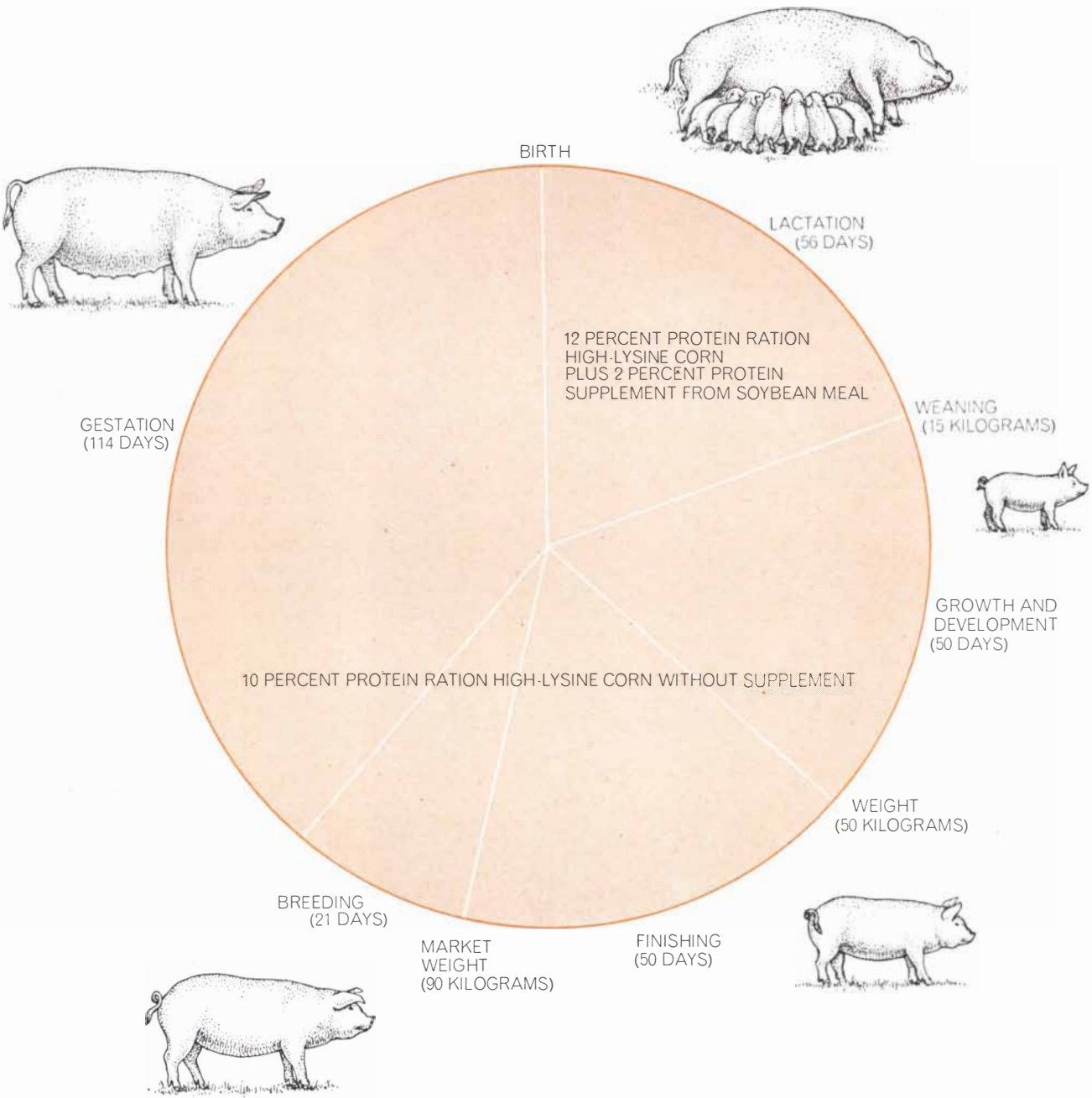


SIMPLIFIED SYSTEM OF PORK PRODUCTION based on high-lysine corn has been developed at the International Center of Tropical Nutrition in Colombia and at several major universities in the U.S. At left is the system based on normal corn; a suitable pro-

meat production with a minimum of supplementation of the corn by imported or costly products. A system using only a small supplement of soybean meal during the growth period and no supplement at all during finishing has worked well, yielding the standard weight achieved by well-fed pigs after 156 days [*see illustration below*]. Similar results have been demonstrated at experiment stations and farms in the U.S.; the economic savings are less striking there, however, because of the ready availability of soybean meal.

Even before the nutrition studies with animals were completed, Alberto G. Pradilla of Colombia recognized the significance of an improved corn protein and began investigations of its characteristics. His work has paralleled the research on the use of high-lysine corn for farm animals, and it has included highly rewarding studies of the value of high-lysine corn as a food for man. The team of physicians working under Pradilla at the University of Valle began in 1967 a series of trials of the new corn for the treatment of malnourished children. They used a diet, previously tested in laboratory animals, that was made up mainly of high-lysine corn with a small supplement of proteins from milk or vegetables. Their first patients were two children, aged five and six, who were suffering from third-degree malnutrition and were under intensive care in the university hospital. The youngsters had been so severely undernourished that their anthropometric development was less than that of a 24-month-old child. On the corn diet they rapidly recovered vital functions, and within 90 days they were showing excellent absorption and retention of nitrogen.

Encouraged by this success, the physi-

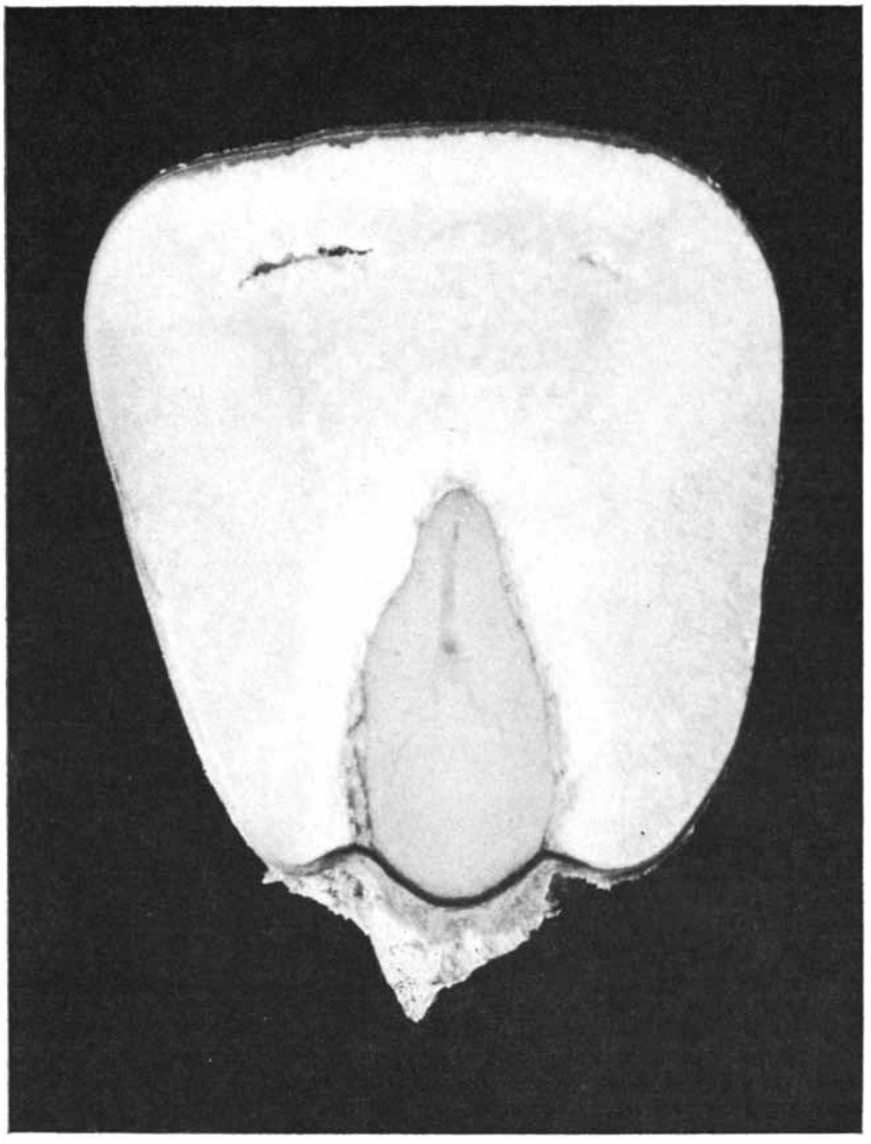


tein supplement must be added throughout the pigs' life cycle. At right is the system based on high-lysine corn; high-lysine corn with supplement is used at one stage and only high-lysine corn is required at all the other stages in the life cycle of the animals.

cians went on to treat children hospitalized for malnutrition with high-lysine corn as the sole source of protein. One of the most remarkable cases of recovery was that of a nine-year-old orphaned girl who had been brought to the hospital by a desperate relative. She was too weak to walk, unresponsive to food, afflicted with severe diarrhea and weighed less than 46 pounds. After a few days of liquid feeding she was able to begin eating high-lysine corn as a solid food. Her response to this diet was so receptive that after 100 days she was discharged from the hospital in good health, with a gain of 11 pounds.

In a country where corn is a familiar and accustomed food, the high-lysine variety is an ideal "prescription" carrying therapeutic values that go beyond its enriched content of protein. The children accepted it readily, without the psychological or physiological resistance that might be offered to a strange diet or "medication." More than 20 malnourished children were treated in the hospital in the series of tests by the university physicians, and they found it to be a very flexible diet, thanks to its acceptability, its superior digestibility over common corn and its essential amino acids. High-lysine corn is now being used successfully for children's diets that, if they were based on normal corn, would be considered well below an optimal level for proper nourishment.

The demonstration of the new corn's ability to "cure" children ill of undernourishment has of course excited considerable interest in its worth as a food among the general population. A food producer in Colombia, the Maizena Company, soon began to market products made from high-lysine corn, notably a precooked baby food offered at relatively low cost. A number of other prepared foods, utilizing high-lysine corn in mixtures or as a precooked corn flour, are being developed. The new corn is still in an exploratory phase as a general food; it has not yet reached the stage where the mass of the population is buying high-lysine corn in the market or growing it for family consumption.



SOFT, FLOURY ENDOSPERM of an opaque-2 kernel is clearly visible in this section. High-lysine corn grinds to a fine flour that calls for new methods of preparation for baking.

Like most radical innovations in the development phase, high-lysine corn still has certain flaws, mostly due to the softness and floury texture of the kernels. It grinds to a fine flour forming a dough that will require new methods of preparation to arrive at the familiar products baked from corn. For industrial processing the only method available to reduce the grain to flour is dry milling. In the field the soft kernel of the growing corn is vulnerable to attack by insects. And the yield by weight is 6 to 10 percent smaller than in normal corn varieties because of the lower density of the kernels. In the present state of the art the user who is least likely to find fault with the new corn is the farmer operating an integrated corn-pig enterprise.

The shortcomings can be corrected. The historic wonders wrought in the breeding of corn have proved it to be such a supple plant that any permutation seems possible. Already there are indications that we can have high-lysine corn with hard, corneous kernels. The breeding work hitherto has focused on the soft-kernel mutants because the distinctive kernel provides a ready, visible marker to indicate the presence of the improved protein. Recently, however, a number of investigators have discovered that hard endosperm not unlike that in common corn occasionally crops up in kernels known to be carrying the opaque-2 gene. These genetic varieties have now been isolated and are being used in efforts to breed high-lysine corn with kernels of normal texture.

Longer-term projects based on the opaque-2 gene are also in view. They call for increasing the total protein content of corn, now averaging about 10 percent, to 12 to 15 percent and for maintaining production yields at the high level characteristic of hybrid corn. If these goals are attained, corn will truly become a "supergrain" carrying an economical answer to the protein needs of peoples in many parts of the world.

We want to be useful ...and even interesting

Kodak

Cheaper packaging

We are pleased to announce a computer-output microfilmer of lower price—purchase or rental—than we have previously offered. At the same time the choice of output formats has broadened, and format has become easy to change.

A computer can tell you much, much more than you want to know unless you keep a firm grip on it. That's where we come in. Look for Kodak in the Yellow Pages under "Microfilming." We show you how to package and distribute the data for convenience, economy, and security. Also sanity.

Two prices for x-rays

In x-ray film for the healing arts, two quite separate kinds of economics enter. The more important of the two reckons in radiation as its coin. The radiation buys information with which to help the patient.

It was on November 8, 1895 that Wilhelm Conrad Röntgen discovered x-rays. Frau Röntgen's hand was the subject of the first human radiograph. Just for fun, of course. Nothing wrong with the hand.

On New Year's Day of 1896 prints of the lady's finger bones were sent out. By spring, amazement had spread from physicists to medical men. Very soon, x-rays became entertainment. The first Nobel Prize ever awarded in physics went to Röntgen.

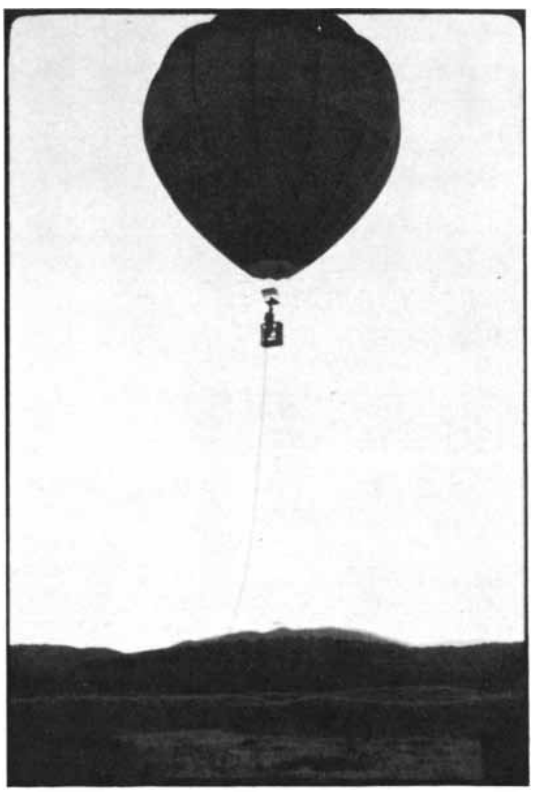
Any kind of photographic plate, film, or paper responded to his amazing rays. Some had to be exposed longer and stronger, but all responded. You merely relaxed.

While the body has no organs designed to feel x-rays, it is not oblivious to them. It took time to learn that. To *that* discovery, Kodak has responded over the years with a long succession of technical stratagems to deliver more information from less radiation. For far less than the radiation price that must have been paid to prove there were bones in Frau Röntgen's hand, more crucial knowledge is bought today, and not just of bones.

But money also counts. In 1906 a 14 x 17-inch glass Kodak plate for an x-ray "sitting" cost the photographer $1. If the client was reasonably well dressed, he might have been wearing a $7 suit and a $3 pair of shoes. By 1918 the plate had dropped in price to 74¢. By 1924, the highly flammable nitrate film that had replaced heavy glass plates had itself been replaced by Kodak safety film, but this was expensive: $1.04 for a 14 x 17-inch sheet.

Today the same size sheet of the newest, vastly more sensitive Kodak x-ray film, designed for machine processing in 90 seconds, can be bought for 84¢. On the other hand, tuition at most first-rate universities has now gone up considerably above $300 a year, and the American Board of Radiology generally confers its certification on an M.D. only after an examination that follows at least four years of radiological residency and one year of joint practice with a radiologist already holding such certificate.

Two academic attitudes toward photography

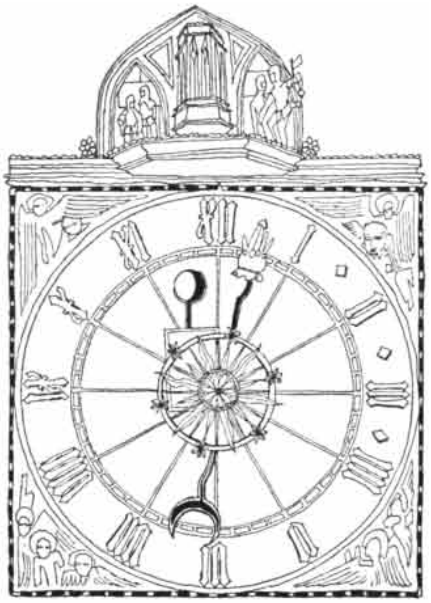


Sent to establish friendly contact for purposes of trade with a band of archeologists encamped in council on what to expect from photography, one of us came back with the above photo of a photographic platform peculiarly useful in their work. What better than a tethered balloon and the right kind of color film for perception of the signs of a site? Like all scientists, archeologists have the problem of explaining themselves beyond the circle that evaluates work by the standards of the disciplines. Surrounding the men and women of science is the larger circle of well-intentioned non-academic humanity who long to understand what it is they are paying for.

The inner circle and the outer circle require very different kinds of photography. The hand that holds the hand-axe in a photograph annoys the scholar by hiding key details from which to identify a culture. But at the expense of keeping that offending hand in the picture, a realtor with a lively mind sees why that particular sharp rock—so like other, undistinguished rocks—shows his ancestors developing the notion of property. A primeval work of art can be photographed for the beauty that the ancient artist gave it. But first it ought to be photographed in a way to record beyond doubt how it stood before the trowels descended and cut for ever the authentic thread to the life of its time.

A small outpost for trading information on archeological photography is maintained by Eastman Kodak Company, Scientific Photography Markets, Rochester, N.Y. 14650.

SCIENCE AND



The Greening of Technology

European science policy will have to be increasingly concerned with social aspirations and goals rather than with simply satisfying the wants of private consumers, according to a reassessment of science policy prepared for the Organization for Economic Cooperation and Development (OECD). Members of the OECD include most of the nations of western Europe. The report, which acknowledges a widespread dissatisfaction with the social costs of rapid economic growth, was prepared by an international committee under the chairmanship of Harvey Brooks of Harvard University.

"Many aspects of developed societies," says the report, "are approaching a condition that may be described as the precursor of saturation, in the sense that things cannot go on growing much longer in some lines without reaching fairly fundamental limits. Indications of saturation are present in total population, pollution of the environment, in the size of [cities], in traffic... even in higher education and perhaps, in the view of some people, the production of new knowledge.... In a society now accustomed to growth in almost all its aspects during the last 300 years, this is something quite new....

"The composition of output is likely to shift from the production of private goods for the market to the production of 'public goods,' i.e., goods and services purchased collectively or supplied by the public sector... education, health care, recreation... clean air and water, parks, public recreation facilities, civic order... housing and urban mass transport.... Productivity in almost all these areas has lagged behind productivity in the manufacture of capital and consumption goods for the private market. Thus science policy will be increasingly concerned with improving productivity in these fields, and with establishing objective criteria of performance."

To bring about a reordering of priorities, the report foresees that the OECD member countries will have to intervene in the working of the market economy in three major ways. "The first is the alteration of the ground rules and incentives under which the market operates by such devices as taxes, subsidies, and judicial actions to internalize [i.e., make explicit] social costs. The second is regulations and standards in relation to such matters as safety, pollution and use of land or other scarce common resources. The third is direct economic activities of government deemed to... warrant supplementing private activities beyond the level that market forces alone would generate. These interventions will affect the allocation of investments and the direction of innovation.... It is now recognized that growth as measured by the market alone does not automatically guarantee the best application of the resources created.... Changes in the quality of life will come only at some cost to other consumption, presumably of the more affluent segments of society.... The control of technology within political systems that encourage competition and pluralism and retain the profit motive represents a major challenge to industry with regard to broadening its motivation."

While emphasizing the need for extending technologies to grapple with pressing social problems, the report urges the continued support of fundamental research, which provides the "basis of new technological developments and for the rational management of technology." Although the report does not refer directly to the controversy over the supersonic transport, it notes that "the emphasis on... environmental management may require that, in the introduction of new technology, [more] fundamental and exploratory applied research to development may be needed than in the past, because the forecasting of side

effects requires more fundamental understanding than is provided by the mere demonstration of technical feasibilities and commercial promise."

The report stresses that if governments do not respond quickly to a growing disenchantment with the kind of technological growth provided by "free" economies, the consequences could be grievous. "The worldwide culture of educated youth, which is deeply concerned with ecological perspectives and is increasingly anti-materialistic, egalitarian, anti-meritocratic, and anti-bureaucratic, could conceivably even adopt anti-rational views and could become much more influential in the next decade than our extrapolations suggest, especially if established methods fail to bring significant improvement. [Such a] failure... could create a... reaction against science [that] could lead to retardation of scientific progress to the point where the world lacked the intellectual tools to cope with the complexity it has created."

Aspirin Inhibits Prostaglandins

Aspirin and certain other anti-inflammatory substances may function, at least in part, by interfering with the synthesis of the prostaglandins. These are fatty acids, only now under intensive investigation, that are widely distributed in the body and that have hormone-like effects on a variety of tissues and life processes. For all of aspirin's long and widespread acceptance as a drug to relieve pain, reduce inflammation and lower fever, its mechanism of action has remained a puzzle. It has come to be considered an "antidefensive" drug that functions by interfering with one or more of the biological substances released by tissues in response to certain stresses.

The prostaglandins are apparently such substances. Several investigators at the Royal College of Surgeons in London have studied the effect of aspirin, related salicylates and the antiinflammatory drug indomethacin on prostaglandins produced in the lung of the guinea pig, in the spleen of the dog and by platelets in the blood of humans. In a series of articles in *Nature New Biology* J. R. Vane, J. B. Smith, A. L. Willis, S. H. Ferreira and S. Moncada report

that the drugs either inhibit or abolish prostaglandin synthesis. The method of inhibition is not known. It may be through competition for the active site of the enzyme that converts arachidonic acid, a precursor, into prostaglandin, Vane suggests; this would conform with the fact that the anti-inflammatory agents in question all have an acidic group and would explain why other kinds of anti-inflammatory agents, such as hydrocortisone, do not inhibit prostaglandin synthesis.

Vane points out that the therapeutic actions of aspirin-like drugs are at least partly consistent with a mode of action involving the prostaglandins, which have been found to induce fever and to be associated with some inflammatory processes. The prostaglandins do not, however, appear to be involved in certain kinds of pain relieved by aspirin (although they can induce headaches). Hence the analgesic action of aspirin remains unexplained. It will probably require experiments on animals or humans to show more directly whether or not prostaglandins are actually the primary agents responsible for, say, headaches or arthritis and therefore whether or not the therapeutic effects of aspirin and indomethacin can be clearly attributed to their inhibition of prostaglandin synthesis.

Protostuff

A meteorite that contains crystalline material older than the solar system has been identified by workers at Case Western Reserve University. The ancient body, called the Allende meteorite, fell near Parral in the Mexican state of Chihuahua on February 8, 1969. Shortly after the fall was discovered specimens of the body were subjected to intensive study by standard optical, petrographic and chemical techniques. The specimens were found to consist of a fine-grained, carbon-rich matrix in which are embedded many chondrules: spherical aggregates of coarser crystals. It was further established that both the matrix and the chondrules are composed predominantly of the mineral olivine (Mg_2SiO_4). On the basis of this early analysis the Allende meteorite was classified as a Type III carbonaceous chondrite.

A more detailed study was subsequently undertaken by H. W. Green II, S. V. Radcliffe and A. H. Heuer at Case Western Reserve to determine the fine structure of the matrix and the chondrules with the aid of a high-voltage transmission electron microscope. It was in the course of this examination that the olivine crystals in several of the chondrules were discovered to have a peculiar fine-scale, black-spot structure, reminiscent of coalesced point defects in ordinary terrestrial crystals. This fine structure, which is lacking in the matrix material, was interpreted as evidence of extensive radiation damage.

By comparing the substructure of the Allende meteorite with that of a comparable meteorite and also of lunar basalts and selected terrestrial rocks, Green, Radcliffe and Heuer were led to conclude that the observed irradiation of the chondrules preceded the cold accretion of matter that took place during the early stages of the formation of the solar system. The radiation, they believe, originated in the protosun. The presence of radiation damage in the chondrule olivine and its absence in the matrix olivine is taken to indicate that the irradiation occurred after the chondrules solidified but before the parent meteorite body was formed. In addition, the investigators point out in *Science*, the evidence obtained from their examination suggests that the Allende meteorite has experienced no significant thermal event and has suffered no significant collisions since its origin. All these observations are in complete agreement with the view that Type III carbonaceous chondrites have not been produced by metamorphic processes from Type II carbonaceous chondrites. "We thus conclude," the investigators write, "that the Allende meteorite consists of virgin planetary material."

Cooling Salts

Air conditioning represents the fastest-growing load borne by the electric utilities, and the one most likely to create "brownouts" on hot summer days. Although the utilities have explored a number of ways to meet peak demands for power, the most direct approach would be to design appliances that would provide a steady load 24 hours a day. Engineers at the University of Pennsylvania have been working on an air conditioner that does just that.

The key is thermal energy storage: during the night and early in the morning the unit "stores cold" in a special salt that freezes at 55 degrees Fahrenheit. When the day warms up and the load on the air conditioner increases, the salt melts, thus absorbing heat and supplementing the basic cooling capacity of the unit. The concept of using salts for thermal storage was pioneered some years ago at the Massachusetts Institute of Technology by Maria Telkes as a way of storing solar energy for house-heating. Now at the University of Pennsylvania she has helped to adapt the heat-storage principle in reverse for air conditioning. According to Steven Freedman, chief engineer on the project, a house that requires a three-ton conventional air conditioner could be adequately cooled by a 1¼-ton unit that runs 24 hours a day on the cold-storage principle. In addition to being 60 percent smaller, the new unit would consume at least 15 percent less power.

Left- and Right-handed Odors

The theory that the shape of a molecule is essentially responsible for its odor has received strong confirmation from two studies of enantiomers, or mirror-image molecules. The theory, originated in 1949 by R. W. Moncrieff and developed by John E. Amoore, conceives of the olfactory receptors as having specifically shaped sites into which molecules of different forms fit, so that the overall configuration of a molecule is more important than its constituent atoms or internal arrangement in determining its odor [see "The Stereochemical Theory of Odor," by John E. Amoore, James W. Johnston, Jr., and Martin Rubin; SCIENTIFIC AMERICAN, February, 1964]. The theory has been much debated, and several other mechanisms for the detection of odor have been proposed, including one that depends on a molecule's vibrational energy levels as exhibited by its infrared spectrum. Clearly one of the sharpest tests of the stereochemical theory would be a comparison of enantiomers. Such molecules are identical in all respects except for their chirality, or handedness. They are related as right and left hands are related and are labeled accordingly: *R* (for "right") and *S* (for *sinister*).

Investigators at the University of California at Davis and at Case Western Reserve University, working independently, studied enantiomers of carvone and several closely related molecules. One enantiomeric form of carvone, *R*-carvone, has a strong odor of spearmint; the other form, *S*-carvone (which is geometrically a mirror image and which in solution rotates the plane of polarization of polarized light in the opposite direction) has the odor of caraway. G. F. Russell and J. I. Hills at Davis started with commercial preparations of the two carvones, purified them and synthesized a number of related compounds, each in the *R* and the *S* form. Each of the compounds was purified by gas chromatography and subjected to odor tests by a panel of judges. In every case the *R* and the *S* forms were

judged to have different odors. The two enantiomers of *trans*-dihydrocarvone, which have geometries similar to carvone's, smelled of spearmint or caraway, depending on handedness. The enantiomers of *cis*-dihydrocarvone, with a slightly different geometry, had a musty or a woody odor, but again the *R* and the *S* forms were clearly distinguishable from each other.

Lester Friedman and John G. Miller, at Case Western Reserve, "interconverted" *R*-carvone and *S*-carvone into their enantiomers and then back again, testing odor along the way. At each step the carvones were found to have identical physical properties except that the *R* and the *S* forms had different odors (*R* spearmint, *S* caraway) and rotated polarized light in opposite directions. Friedman and Miller obtained similar results with other molecules. For example, they showed that the difference between the smell of an orange and the smell of a lemon is attributable to the difference in handedness between *R*-limonene and *S*-limonene respectively. They point out that not all enantiomers smell different; both forms of camphor smell the same. The fact that handedness can so strongly influence odor, however, tends to discredit the vibrational theory and supports a modified form of the stereochemical theory.

Woher der Gegenschein?

A diffuse patch of dim light is sometimes visible near the zenith in the night sky. This is the Gegenschein, or "light of opposition." It was given the name in 1854 by its discoverer, the German astronomer T. J. C. A. Brorsen, because it is located exactly opposite the sun. In the intervening century it has received the attention of many eminent astronomers and has given rise to at least four explanatory hypotheses.

Robert G. Roosen of the National Aeronautics and Space Administration recently summarized the leading hypotheses in *Reviews of Geophysics and Space Physics*. He dismisses all but one of them. One hypothesis was offered by Brorsen some five years after his discovery and was later espoused by others. It is the "gas tail" hypothesis. It visualizes solar radiation sweeping away the hydrogen and helium that leaks out of the earth's atmosphere in a vast plume extending "downstream" from the earth. On this hypothesis the Gegenschein is produced by sunlight that is reflected from the gas tail at a point some 800,000 miles in space. What destroys the hypothesis is the fact that at this relatively short distance the shadow the earth casts into space should produce a dark patch within the brighter Gegenschein. It does not.

Next was the libration-point hypothesis, first suggested in the 1880's and defended at the turn of the century by W. H. Pickering of the Harvard College Observatory. It states that in the dynamic system of the earth and the sun a third point of equilibrium exists a little less than a million miles from the earth. This "libration point" would capture bits of planetary debris, and the sunlight reflected from the material produces the Gegenschein. The fact that kills the first hypothesis does the same to the second. At this distance the earth's shadow should be visible and it is not.

The third hypothesis is a variation on the first: The tail downstream from the earth consists not of gas but of dust, perhaps largely debris generated by the continual bombardment of the moon by meteorites. Again the absence of the earth's shadow is the flaw.

Roosen describes the preferred hypothesis. Since no earth shadow is visible, the reflective material must lie well outside the deep zone covered by the shadow. Might it not be a huge cloud of dust particles traveling in orbit around the sun far beyond the earth? The Italian astronomer G. V. Schiaparelli thought so in 1867, and so today does Roosen. The origin of the distant cloud? It is the debris generated by aeons of collisions in the asteroid belt between Mars and Jupiter.

Cadmium Watch

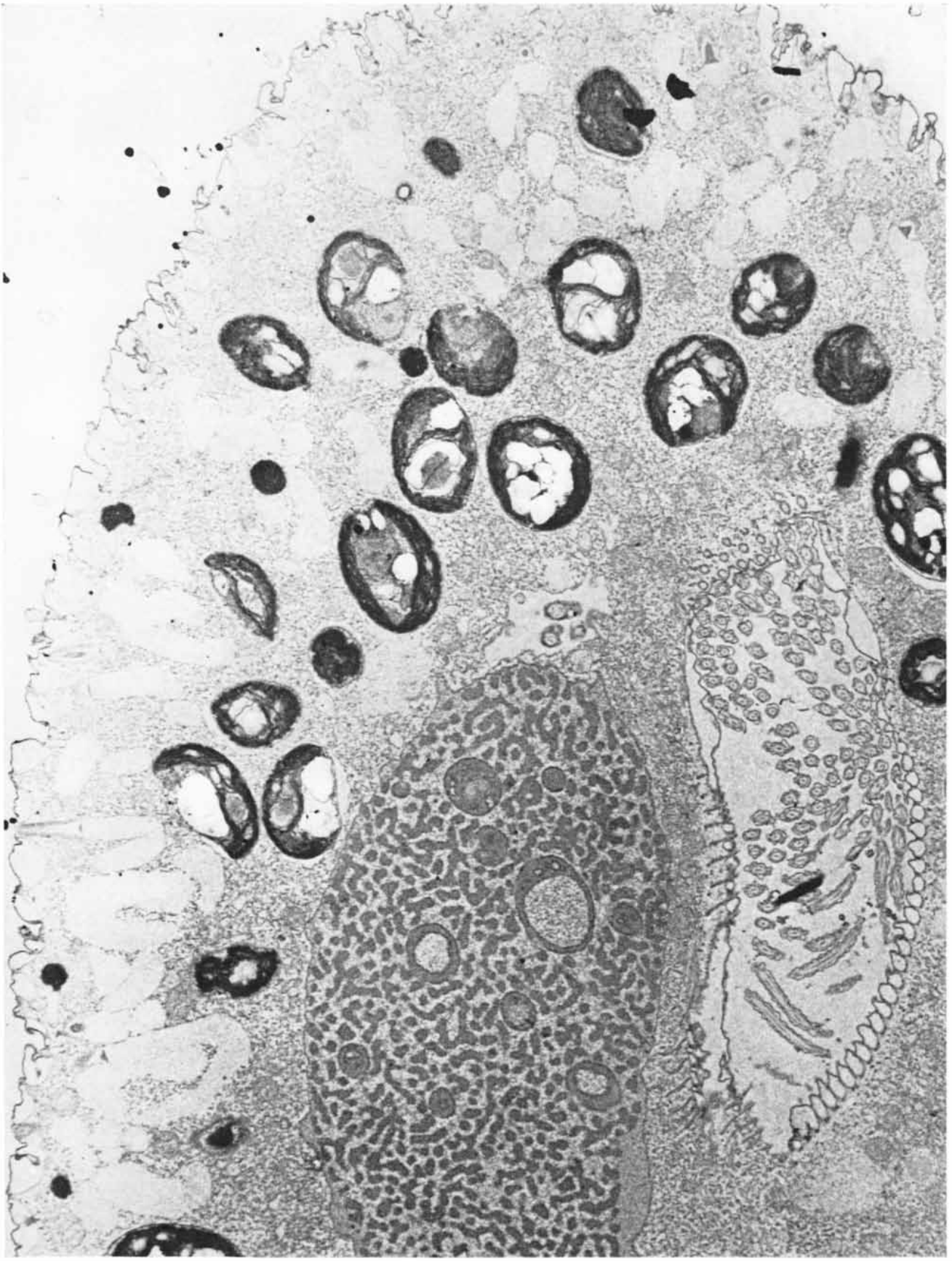
The list of toxic metals in the environment that bear watching is currently being lengthened by the addition of cadmium. Evidence that cadmium in the environment can give rise to "serious intoxication in human beings" is presented in a review prepared by Lars Friberg, Magnus Piscator and Gunnar Nordberg of the Royal Caroline Institute in Stockholm under a contract with the U.S. Environmental Protection Agency. The institute has conducted extensive studies of the effects of metals in the environment. Its review of cadmium was discussed recently at a conference convened for the Environmental Protection Agency by the University of Rochester School of Medicine and Dentistry.

The technological metabolism of cadmium has to do mainly with the manufacture of batteries. The Swedish investigators noted that "the ambient air contains small amounts of cadmium, and in the vicinity of certain factories the concentration can be considerable." Moreover, people are exposed daily to "a substantial amount of cadmium in food." Exposure to the metal in both the industrial environment and the general environment can cause illness, the investigators reported. "Inhalation of cadmium oxide fumes," they write, "can produce acute damage in the lungs in the form of pneumonitis or pulmonary edema," and prolonged exposure to dust or fumes can cause emphysema. They add: "It is the effects due to long-term exposure to low concentrations of cadmium that are of interest. The critical organ is the kidney. Renal tubular dysfunction with proteinuria is a common manifestation. . . . Other systemic effects include anemia and liver dysfunction." Much further work is needed, the investigators said, to clarify the relation between dose and response and to explore more fully the extent, turnover and effects of cadmium in the environment.

Carbon Monoxide Sink

Human activities currently add some 200 million metric tons of carbon monoxide to the atmosphere every year, yet there does not appear to be any buildup of carbon monoxide in the atmosphere as a whole. The principal explanation may lie in the metabolism of microorganisms in the soil. Experiments in which sterile soil failed to take up carbon monoxide whereas most normal soils absorbed it at an average rate of 8.44 milligrams per hour per square meter are described in *Science* by Robert E. Inman, Royal B. Ingersoll and Elaine A. Levy of the Stanford Research Institute. They conclude that soil "must now be considered as a major natural sink for CO that is released into the atmosphere by natural emitters or by the burning of fossil fuels." This sink does not, of course, act quickly enough to affect elevated levels of carbon monoxide in such places as heavily traveled highways.

Inman, Ingersoll and Mrs. Levy estimate that the average capacity of soils in temperate zones for absorbing carbon monoxide is 191.1 metric tons per year per square mile. On that basis they calculate that the capacity of the soil surface of the continental U.S. (2,977,128 square miles) is 569 million metric tons per year, "which is 6.5 times the annual estimated production of CO attributed to the United States and almost three times the estimated worldwide production due to man's activities."



HEREDITARY SYMBIOSIS between photosynthetic green algae of the genus *Chlorella* (*scattered dark ovals*) and a single-celled animal host is characteristic of the species *Paramecium bursaria*, seen magnified 8,000 times in this electron micrograph. Even when the host is kept close to starvation, its guest symbionts satisfy its basic food requirements as long as sunlight is available. The chloroplasts in photosynthetic cells may once have been similar free-living alga-like organisms that eventually became guest symbionts.

SYMBIOSIS AND EVOLUTION

The cells of higher plants and animals have specialized organelles such as chloroplasts and mitochondria. There is increasing reason to believe that these organelles were once independent organisms

by Lynn Margulis

Every form of life on earth—oak tree and elephant, bird and bacterium—shares a common ancestry with every other form; this fact has been conclusively demonstrated by more than a century of evolutionary research. At the same time every living thing belongs primarily to one or another of two groups that are mutually exclusive: organisms with cells that have nuclei and organisms with cells that do not. (An exception is viruses and virus-like particles, but such organisms can reproduce only inside cells.) How can both of these facts be true? Why does so profound a biological schism exist? Ideas put forward and discarded some decades ago hinted at one explanation: Cells without nuclei were the first to evolve. Cells with nuclei, however, are not merely mutant descendants of the older kind of cell. They are the product of a different evolutionary process: a symbiotic union of several cells without nuclei.

The cells of the two classes of organisms are called prokaryotic ("prenuclear") and eukaryotic ("truly nucleated"). The two classes are not equally familiar to us. Most of the forms of life we see—ourselves, trees, pets and the plants and animals that provide our food—are eukaryotes. Each of their cells has a central organelle: a membrane-enclosed nucleus where genetic material is organized into chromosomes. Each has within its cytoplasm several other kinds of organelle. Prokaryotes are far less prominent organisms, although they exist in huge numbers. In the absence of a membrane-enclosed nucleus their genetic material is dispersed throughout their cytoplasm. Such primitive simplicity is characteristic of the blue-green algae and of all the myriad species of bacteria.

The relatedness of living things is fundamental. Organisms as apparently dissimilar as men and molds have almost identical nucleic acids and have similarly identical enzyme systems for utilizing the energy stored in foodstuffs. Their proteins are made up of the same 20 amino acid units. In spite of a bewildering diversity of forms, in these fundamental respects living things are the same. Yet we are left with the equally fundamental discontinuity represented by the two different classes of cells.

Varieties of Symbiosis

Symbiosis can be defined as the living together of two or more organisms in close association. To exclude the many kinds of parasitic relationships known in nature, the term is often restricted to associations that are of mutual advantage to the partners. One frequently cited instance of symbiosis is the partnership sometimes observed between the hermit crab and the sea anemone. The anemone attaches itself to the shell that shelters the crab; this provides its partner with camouflage, and stray bits of the crab's food nourish the anemone. An example that is more pertinent here is the relationship between the leguminous plants and certain free-living soil bacteria. Neither organism can by itself utilize the gaseous nitrogen of the atmosphere. The roots of the plants, however, develop projections known as infection threads that transport the soil bacteria into the root structure. Once present in the cytoplasm of the root cells, the bacteria (transformed into "bacteroids") combine with the host cells to form a specialized tissue: the root nodule. Inert atmospheric nitrogen is utilized by nodule cells as a nutrient. At the same time the nodules manufacture a substance—a pinkish protein known as leghemoglobin—that neither the plant nor the bacteria alone can produce. Because the bacterial symbionts live within the tissue of the plant host the partnership is classified as "endosymbiosis."

Neither of these relationships is necessarily hereditary. The hermit crab will never give rise to the anemone, nor the anemone to the hermit crab. Nor in most instances does a pea or an alfalfa seed contain bacteria; each new generation of plants must establish its own association with a new generation of bacteria. On the other hand, there is one plant—*Psychotria bacteriophila*—that contains the bacterial symbiont in its seed. Thus its offspring inherit not only chromosomes and cytoplasm from the parent plants but bacteria as well. This constitutes hereditary endosymbiosis.

Hereditary symbiosis is surprisingly common. In many instances the host—plant or animal—cannot manufacture its own food and the guest belongs to the family of organisms that can synthesize nutrients by absorbing sunlight. Hosts of this kind are heterotrophs: "other-feeders." Among plants the fungi fit into this group; so do most forms of animal life. Their guests are autotrophs: "self-feeders." The process that nourishes them is the familiar one we call photosynthesis.

An instance of such a relationship is provided by lichens, the characteristically flat, crusty plants that can survive in harshly dry and cold environments. Microscopic study long ago demonstrated that a lichen is a symbiotic partnership between an alga (the autotroph) and a fungus (the heterotroph). Vernon Ahmadjian of Clark University has managed to dissociate the partners that form lichens of the genus *Cladonia,* and he has succeeded in raising the two components independently.

Endosymbiosis has been characterized as swallowing without digesting. One protozoan symbiont—*Paramecium bursaria,* commonly known as the green

paramecium—provides an apt illustration. This protozoon has been studied intensively by Richard Siegel of the University of California at Los Angeles and Stephen Karakashian of the State University of New York at Old Westbury. It is green because numerous photosynthetic green algae inhabit its single cell. The photosynthetic guests, given adequate light, can keep the host alive under near-starvation conditions. When the host is deprived of its guests, it will survive only if extra nutrients are added to its medium. The guests (members of the genus *Chlorella,* a common green alga) will also survive when they are removed from the host.

When the organism is reconstituted in the laboratory by bringing the isolated paramecium and the algae together, an interesting thing happens. Once back inside the host, the algae multiply, but only until the normal, genetically regulated number of algae per paramecium is attained. The multiplication then stops. Should the protozoon encounter free-living *Chlorella,* they are promptly digested. Its own algal partners, however, are totally immune. Somehow the paramecium recognizes its symbiont, although even with the electron microscope it is not easy to see any morphological difference between the free-living *Chlorella* and the symbiotic one.

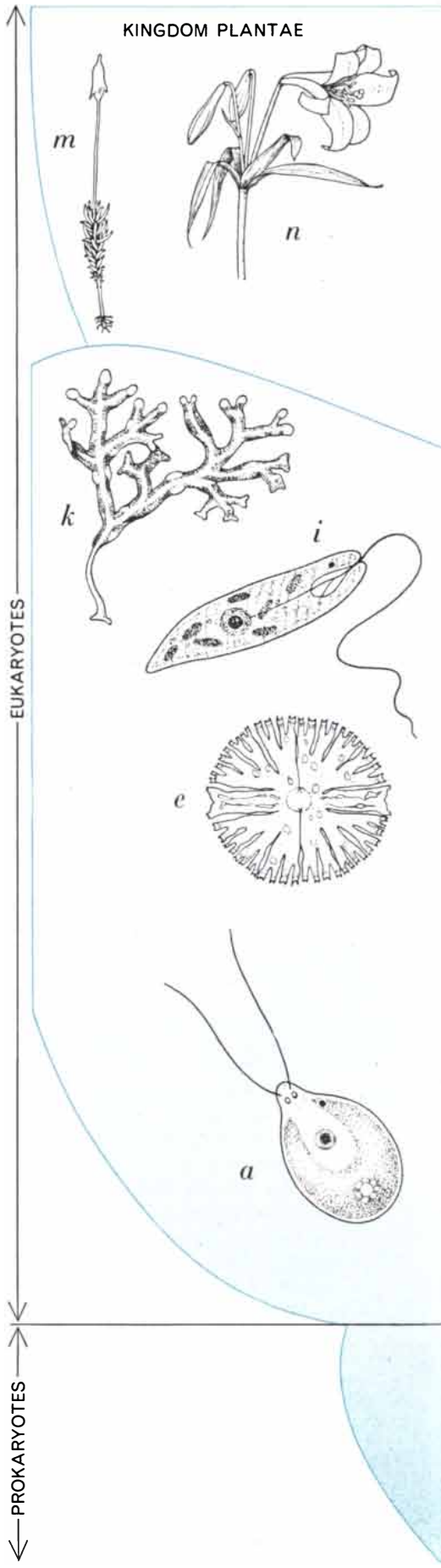
The relationships described thus far involve hosts whose guests all belong to a single species. Far more complex kinds of symbiosis are known. There is one protozoon, for example, that is itself a symbiont and at the same time is the host of three other symbionts. This is the flagellate *Myxotricha paradoxa,* a large, smooth-swimming single-celled organism that seems to be covered with hairlike flagella of various sizes. *Myxotricha* lives in the gut of certain Australian termites; it contributes to the insects' survival by helping them digest the pulverized wood that comprises their food. When *Myxotricha* was first described, it was thought to be just another multiflagellate protozoon with an unusual mode of swimming.

A detailed study by A. V. Grimstone of the University of Cambridge and L. R. Cleveland of the University of Georgia revealed that *Myxotricha* actually had only a few normal flagella at one end. What were mistaken for flagella elsewhere on the organism were spirochetes—a kind of elongated motile bacterium—that were living symbiotically on the surface of the protozoan host. This was not all; each spirochete was associated with another kind of symbiotic bacterium that was also attached to the host's surface, and still a third kind of symbiotic bacterium lived inside *Myxotricha* [*see illustration on page* 52]. As Grimstone and Cleveland have noted, the protozoon "glides along uninterruptedly" through the gut of the termite "at constant speed and usually in a straight line," with its symbiotic spirochetes undulating vigorously.

Organelles of the Eukaryotic Cell

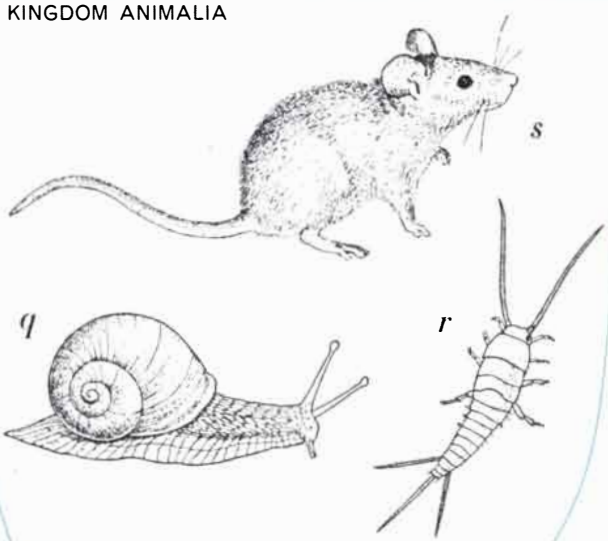
Having seen how many different kinds of independent organism can enter into symbiotic partnerships and how some of these partnerships can be perpetuated on a hereditary basis, we now turn to the eukaryotic cell. When we examine such a cell under the microscope, we see that it contains not only a nucleus but also other organelles. In the eukaryotic cells of a green leaf, for example, there are tiny green chloroplasts, where the chemical events of photosynthesis take place. In the cells of both plants and animals there are mitochondria, where foodstuffs are oxidized to produce ATP (adenosine triphosphate), the universal fuel of biochemical reactions. These are only two of several types of organelle.

Could these organelles have originated as independent organisms? One kind

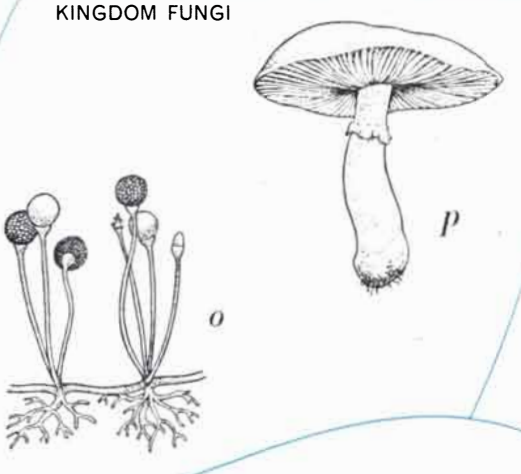


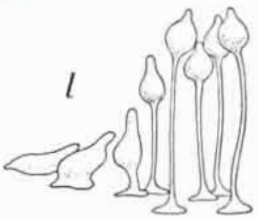
"FIVE-KINGDOM" CLASSIFICATION of terrestrial life, proposed by R. H. Whittaker of Cornell University to solve the dilemma posed by the conventional classification of organisms as either plants or animals, is shown as modified by the author. The life forms comprise two unambiguous and mutually exclusive groups: prokaryotes, the organisms with cells that lack membrane-enclosed nuclei, all within the kingdom Monera, and the eukaryotes, the organisms with truly nucleated cells, which include the populations of the other four. Organisms representative of major phyla are illustrated. In the kingdom Monera these are various bacteria (*left*) and a blue-green alga, *Nostoc* (*right*). In the kingdom Protista are *Chlamydomonas,* one of the chlorophyta (*a*), diatoms (*b*), an amoeba (*c*), a dinoflagellate (*d*), a desmid (*e*), a foraminiferan (*f*), a trypanosome (*g*), a sun animalcule (*h*), a euglena (*i*), a paramecium (*j*), a brown seaweed (*k*) and a cellular slime mold (*l*). The two phyla in the kingdom Plantae, a group nourished by photosynthesis, are represented by a haircap moss (*m*) and a lily (*n*). In the kingdom Fungi, a group characterized by absorptive nutrition, are a bread mold (*o*) and a mushroom (*p*). In the kingdom Animalia, characterized by ingestive nutrition, the representatives are a mollusk (*q*), arthropod (*r*) and chordate (*s*).

KINGDOM ANIMALIA

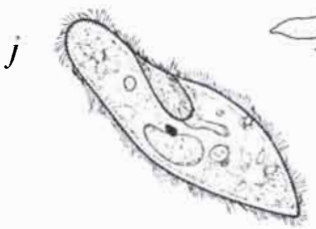


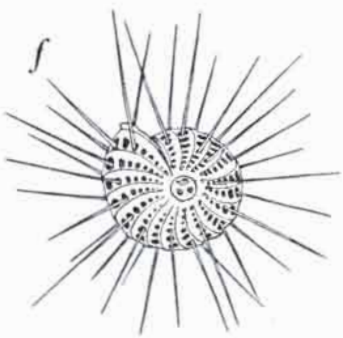
KINGDOM FUNGI

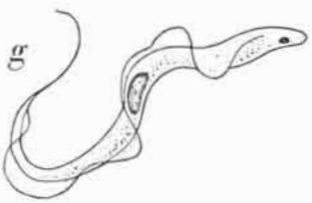


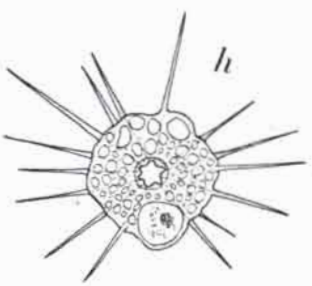


KINGDOM PROTISTA

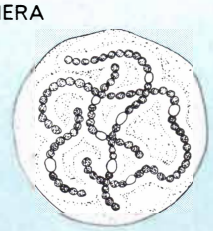








KINGDOM MONERA



of evidence immediately suggests such an origin: the existence of what are known as cytoplasmic genes. When we speak of genes, we usually have in mind the hereditary material—the DNA—in the chromosomes of the cell nucleus. Yet genes are also found outside the nucleus in the cytoplasm, notably in association with chloroplasts and mitochondria.

Chloroplasts belong to a group of organelles collectively known as plastids. Plastids have their own unique DNA—a DNA unrelated to the DNA of the cell nucleus. As has been abundantly demonstrated over the past two decades, DNA is the replicative molecule of the cell. It encodes the synthesis of the proteins required for the doubling of the cell material before cell division. It has also been demonstrated that chloroplasts have their own ribosomes: the bodies where protein is synthesized. The present picture of cellular protein synthesis is that the hereditary information encoded in DNA is transcribed in "messenger" RNA, which then provides the ribosome with the information it needs to link amino acids into a particular protein. In the process each amino acid molecule temporarily combines with a specific molecule of another kind of RNA: "transfer" RNA. Chloroplasts also contain specific transfer RNA's and other components necessary for independent protein synthesis.

Mitochondria also contain DNA that is not related to the DNA of the cell nucleus. The mitochondria in animal cells apparently have only enough DNA and the associated protein-synthesizing machinery to produce a fraction of the structural protein and enzymes needed by these organelles in order to function. Nonetheless, the machinery is there: DNA, messenger RNA, special mitochondrial ribosomes and so forth. The presence of DNA associated with protein synthesis implies that the mitochondria have a functional genetic system.

Here, then, are two organelles of eukaryotic cells that have their own genes and conduct protein synthesis. When one considers that almost all the protein synthesis in the eukaryotic cell is under the direction of nuclear DNA and that the synthesis is accomplished by ribosomes in the cytoplasm external to both the mitochondria and the plastids, it is natural to wonder why these organelles carry duplicate equipment. Does their ability to grow and divide within the cell and to make some of their own protein under the direction of their own genes imply that they were once free-living organisms? A number of investigators have thought so.

When the plastids of eukaryotic algae were studied under the microscope in the 19th century, it was remarked that they resembled certain free-living algae, and it was suggested that they had originated as such algae. A similar origin for mitochondria was proposed in the 1920's by an American physician, J. E. Wallin. On the basis of microscopic observations, of reactions to stains and of assertions (subsequently refuted) that he had grown isolated mitochondria in the laboratory, Wallin maintained that mitochondria were bacteria that had come to live symbiotically within animal cells. In his book *Symbioticism and the Origin of Species* he argued that new species arise as a result of this kind of symbiosis between distantly related organisms. As can happen to people obsessed by a novel concept, Wallin overstated his case and used doubtful data to defend it. His book fell into disrepute.

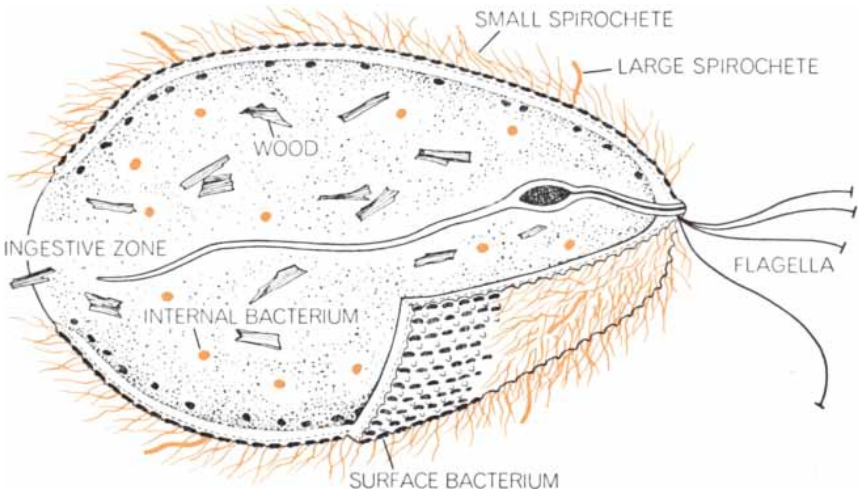
What is known today about the biochemical autonomy of mitochondria goes a long way toward rehabilitating Wallin's basic concept. It now seems certain that mitochondria were once free-living bacteria that over a long period of time established a hereditary symbiosis with ancestral hosts that ultimately evolved into animal cells, plant cells and cells that fit neither of these categories. The same history evidently holds true for plastids, which were originally free-living algae. I believe that still a third group of organelles, the flagella and cilia, became associated with the eukaryotic cell in much the same way.

Flagella and Cilia

Flagella and cilia are really the same. If these hairlike cell projections are long and few, they are called flagella; if they are short and many, they are called cilia. Their motion propels the cell through its medium or, if the cell is fixed in place, moves things past it. In the tissues of higher animals some flagella and cilia have been drastically modified to serve other functions. The light receptors in the eye of vertebrates are such structures. So are the smell receptors of vertebrates. Among prokaryotes the analogous structures are much simpler. They are small, single-stranded and consist of a protein called flagellin.

The flagella and cilia of eukaryotic cells are much larger than those of prokaryotes. Their basic structure is strikingly uniform, whether they come from the sperm of a fern or the nostril of a mouse. Seen in cross section, each consists of a circle of paired microtubules surrounding one centrally located pair. If the structure is motile, there are always two microtubules in the middle and always nine more pairs surrounding them; the pattern is known as the "9 + 2 array" [*see illustrations on page 54*]. Microtubules from any kind of eukaryotic flagella and cilia are composed of related proteins called tubulin.

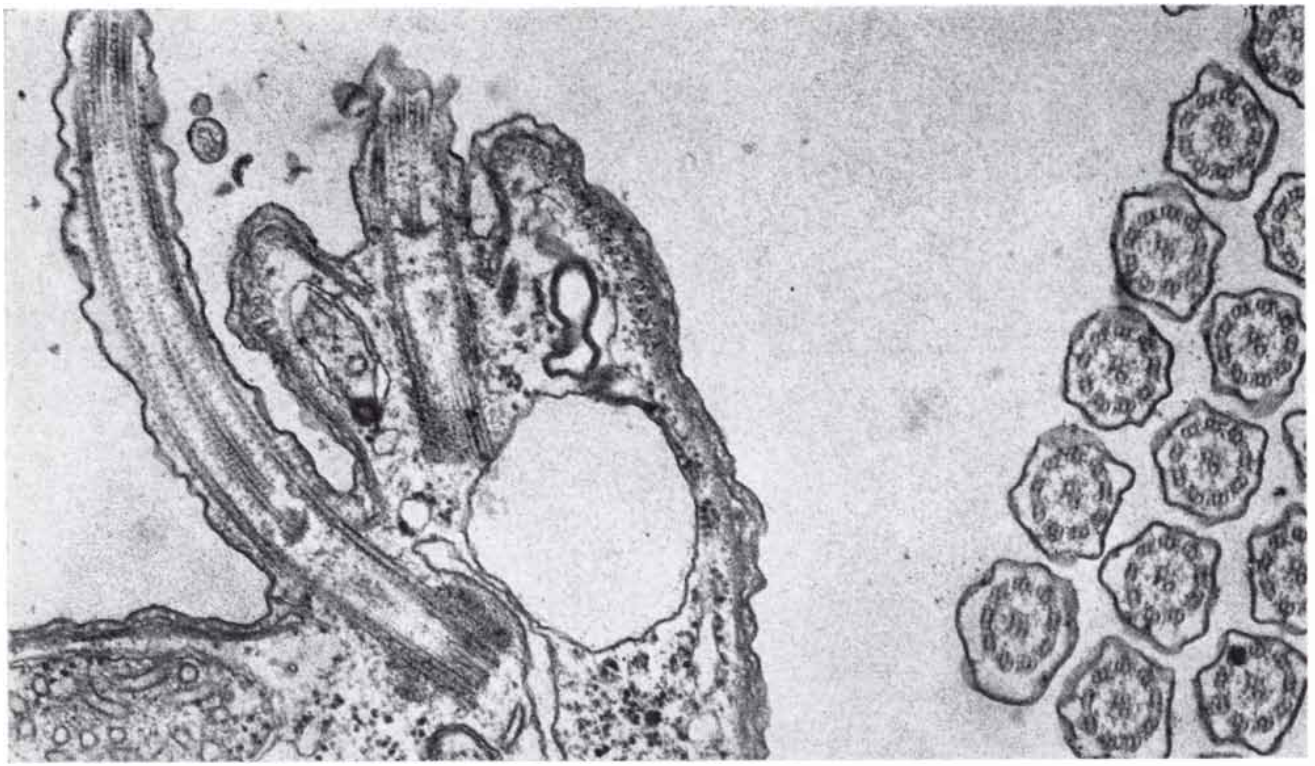
At the base of every eukaryotic flagellum and cilium is a distinct microtubular structure: the basal body. The architecture of the basal body is identical with that of the centriole, a structure found



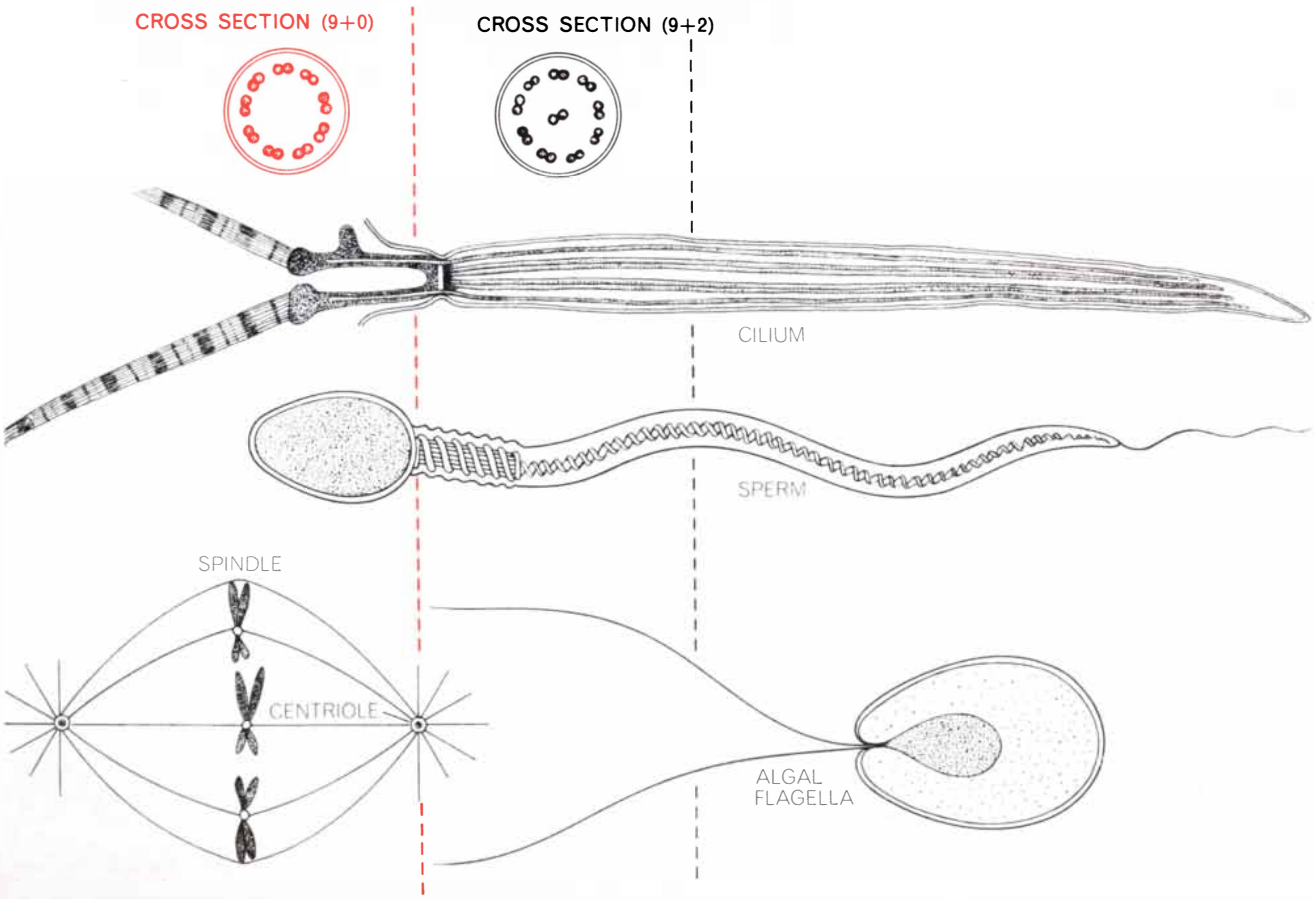
COMPLEX SYMBIONT, the protozoon *Myxotricha paradoxa*, lives as a guest in the gut of certain Australian termites and plays host to three symbionts of its own. These are surface bacteria of the spirochete group (*color*), which observers first mistook for flagella, other surface bacteria (*black*) and still other bacteria (*color*) that live inside the protozoon.

SURFACE OF *MYXOTRICHA* appears at the bottom in transverse section in this electron micrograph by A. V. Grimstone of the University of Cambridge. In the "hollow" to the left of each surface "peak" lies one of the bacterium guests of the protozoan host. Two symbiotic spirochetes are visible at right with their basal ends attached to the host membrane. Other spirochetes whose attachments are not in the plane of focus are partially visible elsewhere in the micrograph (*top*). The theory proposing that eukaryotic cells are the products of similar symbiotic relationships suggests that the first symbionts were free-living bacterium-like prokaryotes.

PROKARYOTIC GUESTS, identifiable by their array of concentric photosynthetic membranes as the blue-green alga *Cyanocyta*, are enlarged 15,000 times in this electron micrograph by William T. Hall of the National Institutes of Health. They are inside protozoan hosts of the species *Cyanophora paradoxa*. Similar hereditary symbioses between various photosynthetic alga-like prokaryotes and large, more advanced eukaryotic hosts from the kingdom Protista is suggested as the step leading to evolution of the plant kingdom.



STRUCTURE OF FLAGELLA is shown in transverse section (*right*) and longitudinal section (*left*) in an electron micrograph by R. D. Allen of the University of Hawaii. In the part of the flagellum extending beyond the basal body a circle of paired microtubules surrounds a central pair in what is known as a "9 + 2 array." In the basal body the central pair of microtubules is absent and the array is "9 + 0." Such organelles are found only among the eukaryotes and may originally have been free-living cells.



MICROTUBULES comprise a variety of structures, including the motile flagella of certain algae (*bottom right*) and of sperm, the cilia of tracheal membrane and the centrioles and the spindle structure that mediates halving of the nucleus in mitotic division.

at opposite poles of the eukaryotic cell nucleus. Centrioles come into particular prominence during mitosis, the process by which eukaryotic cells divide. (Centrioles are found in nearly all animal cells and in the cells of many eukaryotic algae but not in certain fungi and in most higher plants).

The structural array of the basal body and the centriole is "9 + 0": the central pair of microtubules is absent. In cells that possess mitotic centrioles the centrioles left over from earlier cell divisions often grow projections that become flagella or cilia as the new cell differentiates. Thus not only are basal bodies and centrioles identical in structural pattern but also centrioles can become basal bodies. Moreover, the mitotic spindle, the characteristic diamond-shaped structure that lies between the centrioles during cell division, is an array of microtubules composed of tubulin.

A further finding requires that we now ask two fundamental questions. When the plant alkaloid colchicine is added to tubulin, derived either from flagella or cilia or from spindles, the alkaloid is bound to the protein. The reaction is characteristic of tubulin from the cells of all animals and all eukaryotic plants, but it has never been observed with the flagellin from prokaryotic cells. Nor, for that matter, have microtubules ever been observed in either bacteria or blue-green algae.

The first question is this: What differentiates animals from plants? At the macroscopic level the differences are obvious; for example, most animals move around in order to feed themselves, whereas most plants stand still and nourish themselves by photosynthesis. At the microscopic level distinctions of this kind become meaningless. Many kinds of single-celled organism sometimes nourish themselves by photosynthesis and at other times swim about ingesting food particles. Some organisms crawl like an amoeba at one stage in their development but later stop, sprout stems and disperse a new generation in the form of spores. Further examples are almost innumerable.

Generations of biologists have been troubled by the need to force such organisms into the plant or animal kingdom. A far less ambiguous dichotomy is the division between prokaryotes and eukaryotes. Notable dissenters from the plant-animal classification are Herbert F. Copeland of Sacramento State College, G. Evelyn Hutchinson of Yale University and most recently R. H. Whittaker of Cornell University. In what follows I have modified Whittaker's "five-kingdom" classification, which takes the fundamental prokaryote-eukaryote dichotomy fully into account [*see illustration on pages 50 and 51*]. The answer to the first question, then, is that there are not just two basic kinds of organism but five.

This brings us to the second question: How did five kingdoms arise? I have already suggested that eukaryotic cells, which are characteristic of all higher forms of life, came into existence through an evolutionary advance of a kind fundamentally different from discrete mutation. Specific answers to the second question will appear in the following hypothetical reconstruction of the origin of eukaryotic cells. The reconstruction traces the rise of the more advanced four of Whittaker's five kingdoms from their origin in the least advanced one. That kingdom is the kingdom Monera: the prokaryotic single-celled organisms that were the first living things to evolve on the earth. The reader should be warned that my presentation of the theory here is necessarily brief and oversimplified.

The First Cells

All life on the earth is believed to have originated more than three billion years ago during Lower Precambrian times in the form of bacterium-like prokaryotic cells. At that time there was no free oxygen in the atmosphere. The cells that arose were fermenting cells; their food consisted of organic matter that had been produced earlier by the action of various abiotic processes. Under pressures of natural selection directly related to the depletion of this stock of abiotic nutrients, there arose among the first fermenting bacteria many metabolic traits that are still observable among bacteria living today. These traits include the ability to ferment many different carbohydrates, to incorporate atmospheric carbon dioxide directly into reduced metabolic compounds, to reduce sulfate to hydrogen sulfide as a by-product of fermentation, and so on.

As the ammonia available in some parts of the environment became depleted, certain bacteria evolved metabolic pathways that could "fix" atmospheric nitrogen into amino acids. Other fermenters developed into highly motile organisms that foreshadowed such highly mobile living bacteria as spirochetes. All these fermenting bacteria were "obligate anaerobes," that is, for them oxygen was a powerful poison. Through various detoxification mechanisms the fermenters were able to dispose of the small amount of this deadly element present in the environment as a result of abiotic processes. Finally, many if not all of the various fermenting bacteria were equipped with well-developed systems for the repair of DNA. Such systems were necessary to counteract the damage done by ultraviolet radiation, which at that time was intense because there was no ozone (O_3) in the atmosphere to filter it out.

All these bacteria were heterotrophs; they had not evolved the photosynthetic mechanisms that would have enabled them to nourish themselves in the absence of abiotic organic compounds. In time some of them developed metabolic pathways that led to the synthesis of the compounds known as porphyrins. It is a purely fortuitous property of porphyrins that they absorb radiation at the visible wavelengths; nonetheless, this property was eventually put to use by the evolution of bacterial photosynthesis. The process of photosynthesis requires a source of hydrogen. Bacteria can utilize such inorganic substances as hydrogen sulfide and gaseous hydrogen as well as various organic compounds of the kind that would have been present in the environment as by-products of fermentation. These first anaerobic photosynthesizers appeared in Lower Precambrian times.

When the new photosynthetic bacteria became well established, a process that may have taken millions of years, a second kind of photosynthesis was able to make its appearance. In the second process the uptake of hydrogen was accomplished by the splitting of water molecules; as a result increasing quantities of lethal free oxygen entered the atmosphere as a waste product. The evolution of this mode of photosynthesis led to the appearance of the blue-green algae, the first organisms on the earth that were adapted to the presence of free oxygen. Since they were active photosynthesizers of the newer type, they accelerated the increase in atmospheric oxygen.

The blue-green algae, whose Precambrian success is attested by the massive calcium-rich rock formations they left behind, presented a profound threat to all other forms of life. The other organisms were forced to adapt or perish. Some of the anaerobes adapted simply by retreating into the oxygen-free muds where their fellows are found today. Others developed new mechanisms of oxygen detoxification; still others, it is safe to assume, merely disappeared. In any case, one result of the success of the blue-green algae was the evolution of new kinds of bacteria that utilized free

oxygen in their metabolic processes: aerobic respirers, oxidizers of sulfide and ammonia, and the like. As atmospheric oxygen continued to accumulate, the stage was set for the initial appearance of eukaryotic cells.

The First Eukaryote

The first advanced cell came into existence when some kind of host, perhaps a fermenting bacterium, acquired as symbiotic partners a number of smaller oxygen-respiring bacteria. As atmospheric oxygen continued to increase, selection pressure would have favored such a symbiosis. Eventually the small aerobic bacteria became the hereditary guests of their hosts; these were the first mitochondria. The host symbionts, in turn, evolved in the direction of amoebas, so that a new population of large aerobic cells evolved and faced the problem of finding nutrients.

In due course the partners were aided in their quest for food: a second group of symbionts, flagellum-like bacteria comparable to modern spirochetes, attached themselves to the host's surface and greatly increased its motility. If this hypothetical triple partnership begins to resemble the termite symbiont *Myxotricha*, it is with good reason; I believe that just such a *Myxotricha*-like symbiotic association, formed in Precambrian times, was a universal ancestor to all eukaryotic organisms. With the appearance of this supercell the kingdom Monera gives rise, in a manner consistent with Whittaker's taxonomic system, to the kingdom Protista.

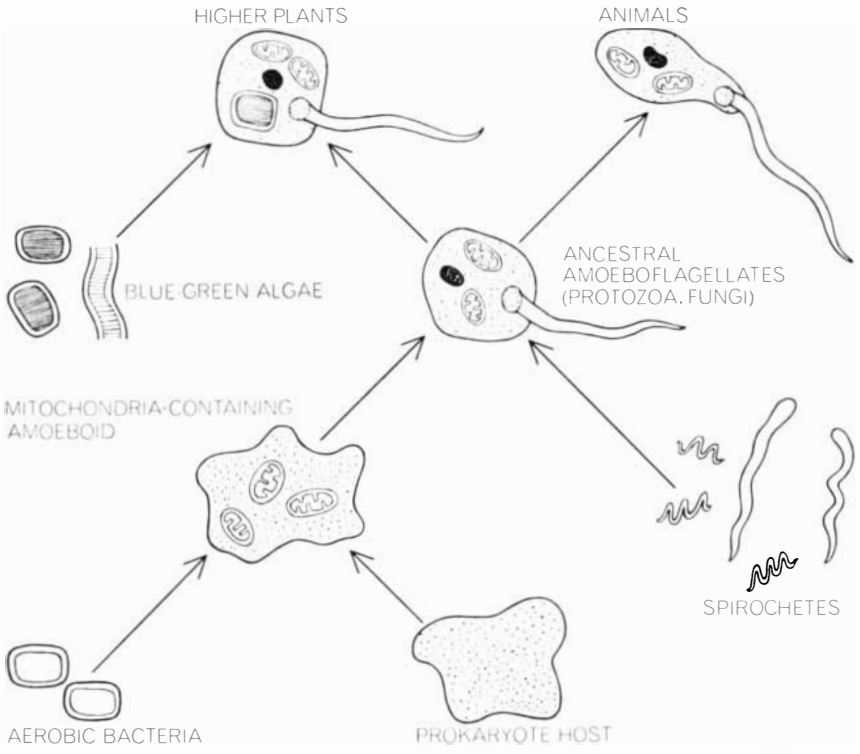
The internal guests, then, served as mitochondria and the external ones as flagella. The spirochete-like guests, however, slowly evolved another role. The specialized basal body of the flagellum and its associated microtubules came to serve the additional function of mediating the process of cell division. Respectively the centriole and the mitotic spindle, they were responsible for dividing the parent cell's genes evenly between daughter cells.

Mitotic cell division was the crucial genetic step toward further evolutionary advance. One would not expect it to have developed in a straight-line manner, starting with no mitosis and concluding with perfect mitosis. There must have been numerous dead ends, variations and byways. Evidence of just such uncertain gradualism is found today among the lower eukaryotes, for example the slime molds, the yellow-green and golden-yellow algae, the euglenids, the slime-net amoebas and others. Many of their mitotic arrangements are unconventional. The perfection of mitosis may have occupied as much as a billion years of Precambrian time.

Mitosis, however, was the key to the future. Without mitosis there could be no meiosis, the type of cell division that gives rise to eggs and sperm. There could be no complex multicellular organisms and no natural selection along Mendelian genetic lines. As mitosis was perfected the kingdom Protista gave rise to three other new kingdoms.

Plant-like protists probably appeared several times through symbiotic unions between free-living, autotrophic prokaryote blue-green algae and various heterotrophic eukaryote protists. After much modification the guest algae developed into those key organelles of the plant kingdom, the photosynthetic plastids. Some of the original symbiotic organisms are represented today by the eukaryotic algae that eventually evolved into the ancestors of the plant kingdom. Both algae with nucleated cells and higher plants have of course evolved a great deal since they first acquired photosynthetic guest plastids more than half a billion years ago. Their evolutionary progress, however, involves neither the origin nor any fundamental modification of the photosynthetic process. This heritage from their anaerobic prokaryote ancestors they received fully formed at the close of the Precambrian.

The group of organisms that we know as the fungi—molds, mushrooms, yeasts and the like—are also thought to derive directly from protists that relinquished flagellar motility in exchange for mitosis. This suggestion is consistent with Whittaker's classification. He splits the fungi from the plant kingdom and recognizes that these fundamentally different organisms deserve a domain of their own. The evolution of the animal kingdom, in turn, is considered a straight-line consequence of natural selection acting on the multicellular, sexually reproductive organisms that, like the fungi, did not happen to play host to plastids in Upper Precambrian times.



SYMBIOSIS THEORY is summarized in the three steps illustrated here. Union between two members of the kingdom Monera, a newly evolved aerobic bacterium (*bottom left*) and a larger host, possibly a fermenting bacterium (*bottom right*), brought into existence an amoeboid-like protist whose several guests became mitochondria. A second hereditary symbiosis, joining the amoeboid to a bacterium of the spirochete group (*center right*), brought into being an ancestral "amoeboflagellate" that was the direct forebear of two kingdoms: Fungi and Animalia. When the same amoeboflagellate went on to form another relationship, with algae that became plastids, the fifth kingdom, Plantae, was founded.

Testing the Hypothesis

Compared with what had gone before, however, all this seems to be virtually modern history. It is more pertinent at this juncture to see if the theory of

FREE-LIVING MONERAN, a bacterium of the spirochete group, is seen magnified 55,000 times in this electron micrograph. It is an anaerobic bacterium found in the human mouth. Organisms like this may have given rise to the eukaryotic flagella through symbiosis.

eukaryotic-cell origin through hereditary symbiosis offers useful answers to further outstanding questions.

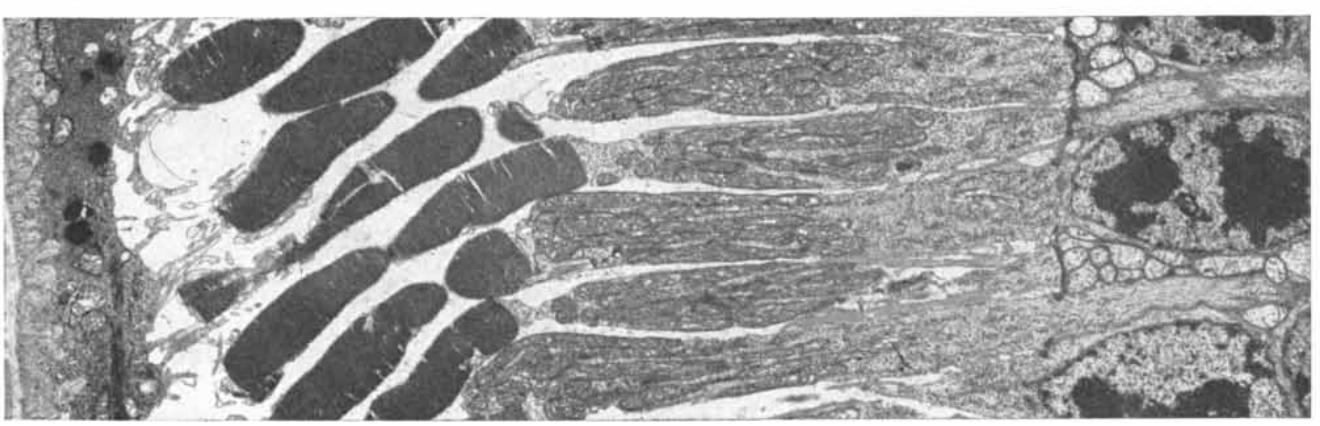
Why are there genes outside cell nuclei? Some cytoplasmic genes may have arisen in other ways, but the symbiosis theory holds that the genes associated with chloroplasts and mitochondria demonstrate that these two kinds of organelle were once free-living organisms.

Why does evidence for photosynthesis appear in Middle Precambrian times, even though no higher plants appear in the fossil record until a mere 600 million years ago? The theory proposes that the higher plants are the result of a symbiosis between animal-like hosts and photosynthetic blue-green-alga-like guests whose partnership could not have evolved until relatively recent times, when mitosis had been perfected.

Why should there be any connection between, on the one hand, the basal body and the flagellum and, on the other, the centriole and the mitotic spindle? The proposal is that the original free-living organism that once accounted only for the function of motility was ancestral to the organelles that came to mediate the equal partition of genetic material between daughter cells during mitosis.

Obviously many other questions remain to be answered. Can the synthesis of DNA and of messenger RNA be detected in association with the reproduction of the basal body and the centriole? Can evidence be found of a unique protein-synthesis system associated with these bodies? Without such evidence the case for these organelles having once been free-living organisms is weak. How and when did meiosis evolve from mitosis? Which organisms were the initial hosts to the guest bacteria that became mitochondria? Were guest plastids of different kinds—red, brown, golden-yellow—acquired independently by the various kinds of eukaryotic algae? One related question is profoundly social. Can botanists, invertebrate zoologists and microbiologists, with their widely different backgrounds, agree on a single classification and a consistent evolutionary scheme for the lower organisms?

Conclusive proof that the symbiosis theory is correct demands experiment. The symbiotic partners will have to be separated, grown independently and then brought back into the same partnership. No organelle of a eukaryotic cell has yet been cultivated outside the cell. The function of a theory, however, is to make reasonable predictions that can be proved or disproved. The predictions of the symbiosis theory are clear.



SPECIALIZED FLAGELLA, shown enlarged 3,900 times in an electron micrograph by Toichiro Kuwabara of the Harvard Medical School, are visual receptors in the retina of a rabbit. The darker structures at left are the outer segments of the visual receptors.

A two-inch empty pipe
can carry 230,000 telephone conversations.

The pipe is no bigger than your wrist.

Yet what really makes it news is that there's absolutely nothing inside.

Except room for 230,000 simultaneous telephone conversations.

In the years to come, millimeter waveguide pipe will be buried four feet underground. In a larger cradling pipe to give it protection and support.

It'll also have its own amplifying system about every 20 miles. So your voice will stay loud and clear.

Even after 3,000 miles.

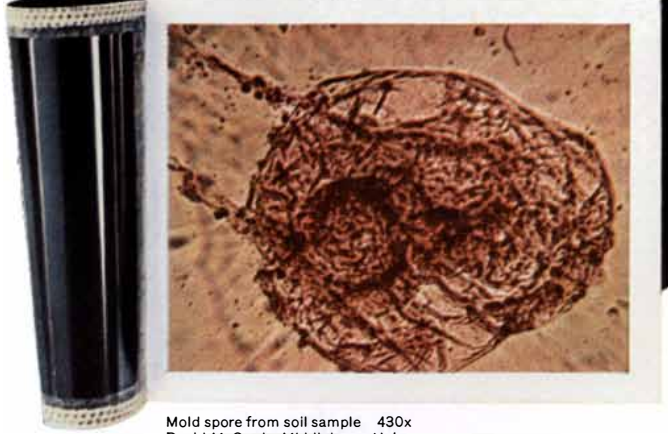
Yet this little pipe is capable of carrying a lot more than just conversations.

It can also carry TV shows. Picturephone® pictures. Electrocardiograms. And data between thousands of computers.

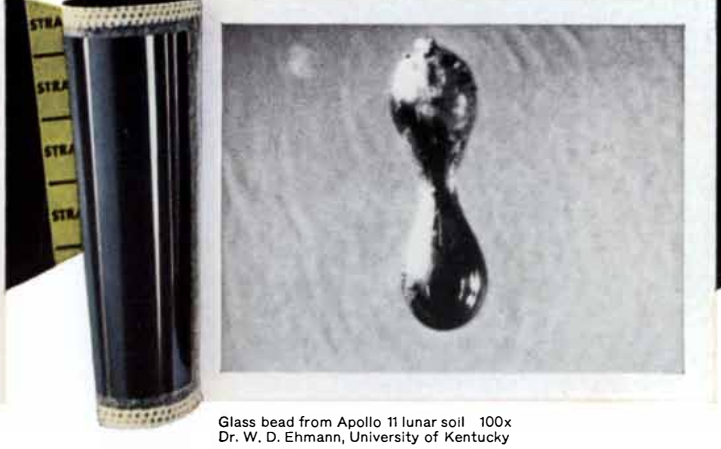
All at once.

The American Telephone and Telegraph Company and your local Bell Company are always looking for new ways to improve your telephone service.

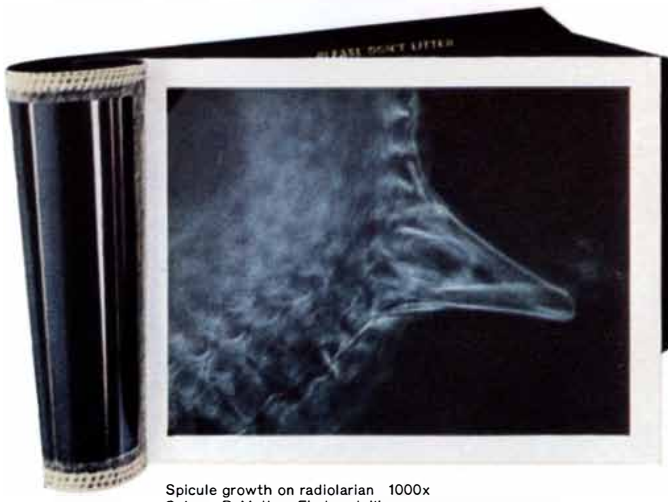
Sometimes that means developing a better way to use two inches of empty space.



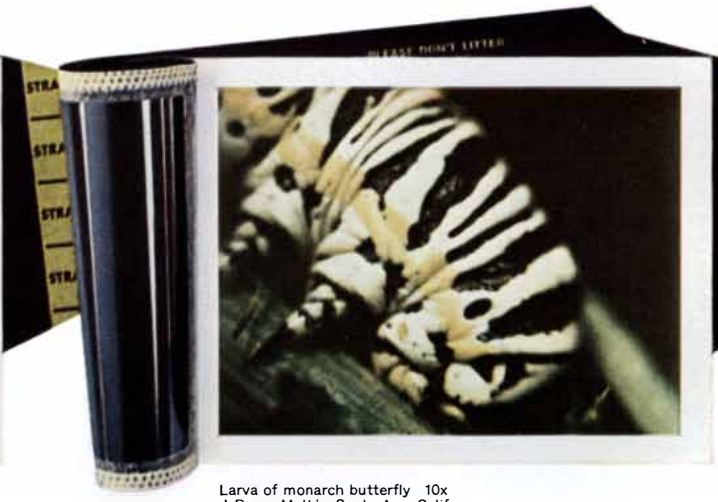
Mold spore from soil sample 430x
David McCurdy, Middletown, N.J.



Glass bead from Apollo 11 lunar soil 100x
Dr. W. D. Ehmann, University of Kentucky

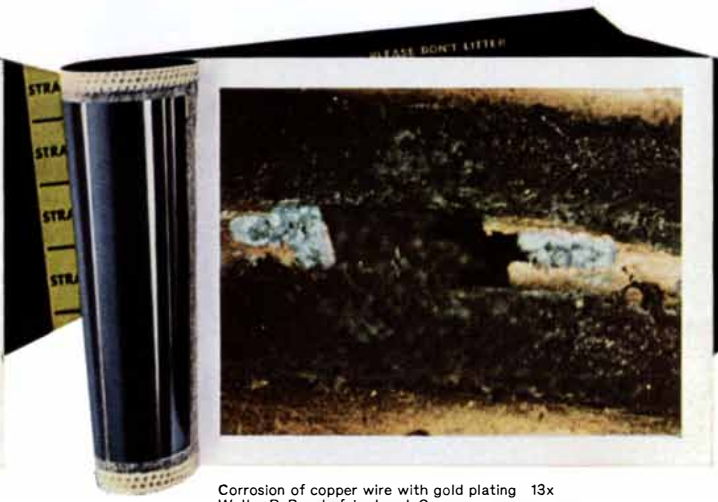


Spicule growth on radiolarian 1000x
Selwyn R. Mather, Elmhurst, Ill.



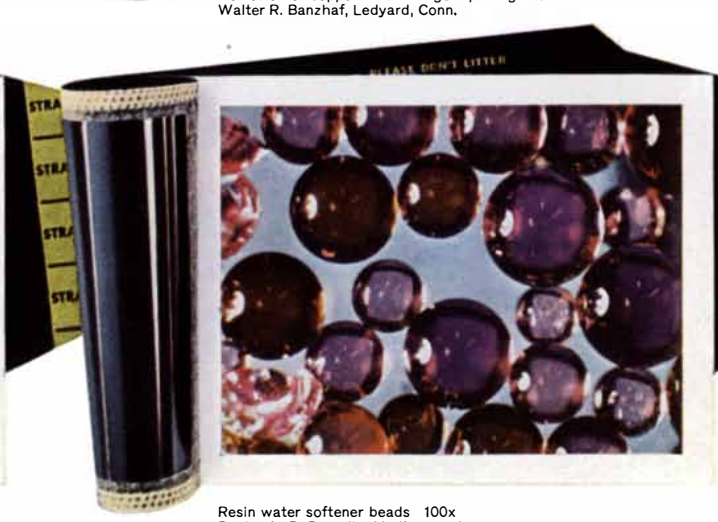
Larva of monarch butterfly 10x
J. Roger Matkin, Santa Ana, Calif.

Freshwater copepod cyclops 80x
Robert J. Western, Kailua, Hawaii



Corrosion of copper wire with gold plating 13x
Walter R. Banzhaf, Ledyard, Conn.

Bone section from 8000-year-old goat 35x
Dr. Isabella M. Drew, Sackler Lab., Columbia University



Resin water softener beads 100x
Benjamin B. Bonadio, Madison, Ind.

All it takes is a microscope, a minute, and less than $70.

You're looking at some winners in a nationwide photomicroscopy contest conducted by Polaroid Corporation.

All entries were taken with the Polaroid ED-10 Land Instrument Camera. Despite its low cost, you can see that it produces pictures as high in quality as those taken with the most expensive microscope cameras. Because the quality depends on the microscope's own optics.

The ED-10 fits virtually any microscope. Whether monocular, binocular or trinocular. And can be attached in moments. A universal adapter attaches to the eyepiece, and you simply slide the camera over the adapter.

You don't even have to know photography to use it. All you do is frame and focus the specimen in the microscope, slip on the focusing tube to refocus in the camera's film plane, replace the tube with the camera, and snap the picture.

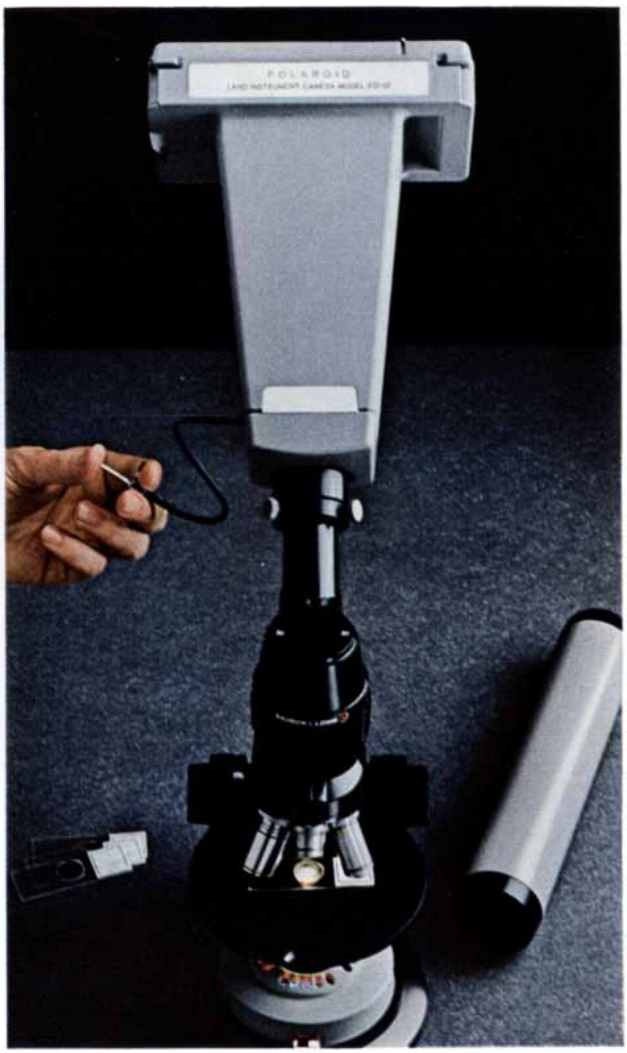
And you don't wait hours or days for results. All it takes is a minute for a full-color, finished photomicrograph and only 15 seconds for black-and-white.

You use Polaroid Land pack film, either Polacolor Type 108, or black-and-white Type 107 (3000 speed). Both are available almost everywhere film is sold.

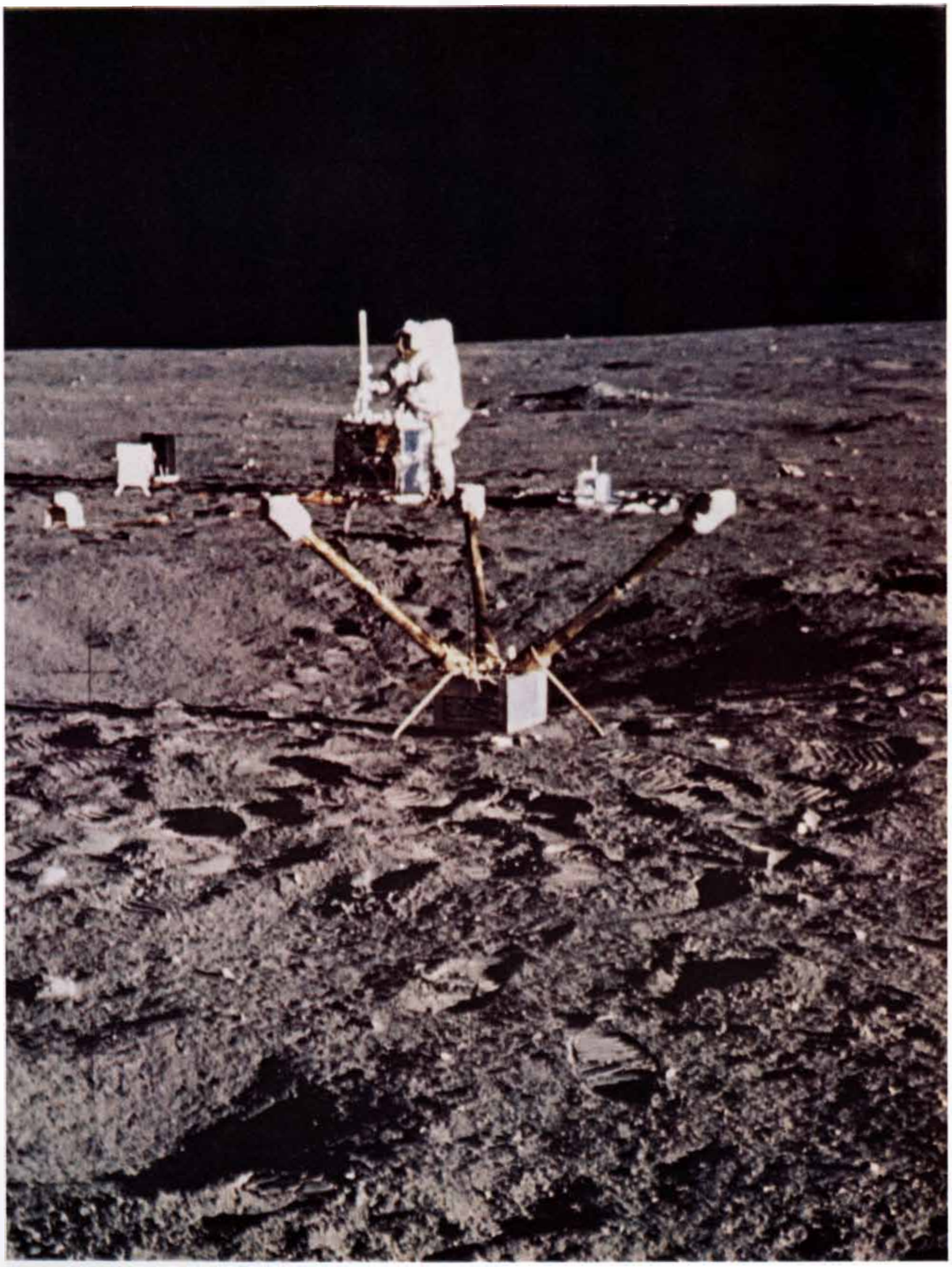
The ED-10's body is molded of high-impact ABS material. So it's rugged, lightweight and remarkably easy to handle.

You can buy it at scientific supply houses, microscope dealers and many camera stores. But if you'd like more information first, write to Polaroid Corporation, Dept. 57-213, 549 Technology Square, Cambridge, Mass. 02139. In Canada, 350 Carlingview Drive, Rexdale, Ont.

It's well worth looking into. "Polaroid" and "Polacolor"®



The Polaroid Land Instrument Camera.



APOLLO 12 SURFACE MAGNETOMETER was photographed by astronaut Alan L. Bean just after he finished deploying it near the *Apollo 12* landing site in Oceanus Procellarum. Astronaut Charles Conrad, Jr., can be seen in the background aiming the central-station S-band antenna toward the earth. The magnetometer was isolated from other scientific instruments in order to avoid measuring their magnetic fields. The device operates continuously during lunar day and night on power from a nearby nuclear generator.

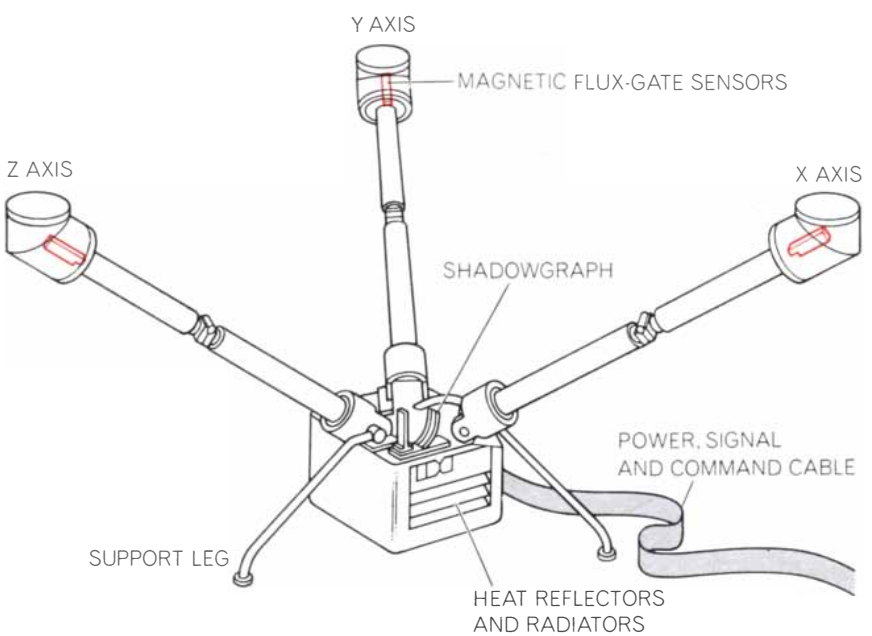
THE MAGNETISM OF THE MOON

Measurements sent to the earth by magnetometers placed on the lunar surface by the Apollo astronauts reveal much about the history and present physical state of the moon

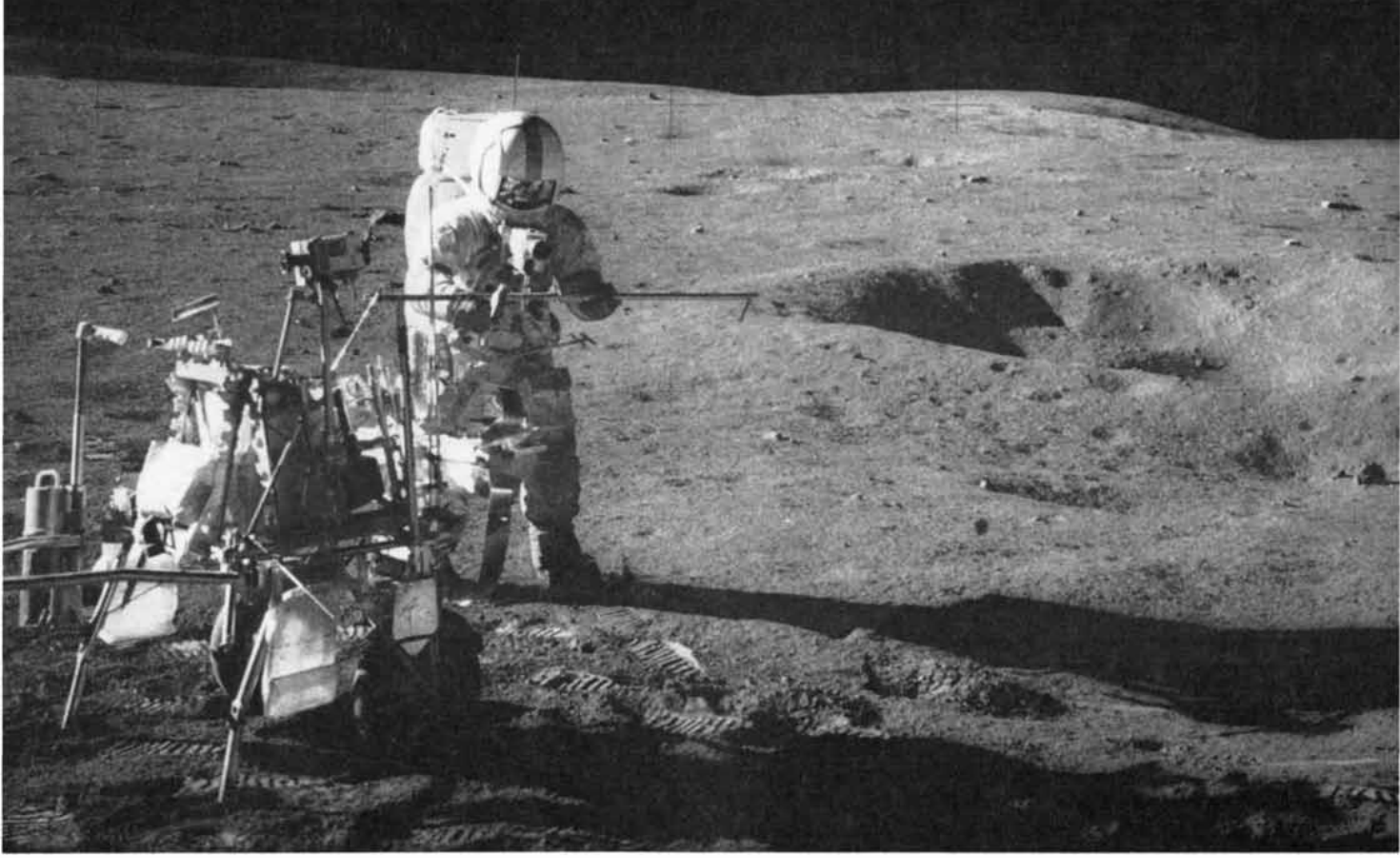
by Palmer Dyal and Curtis W. Parkin

Magnetometers placed on the moon by the astronauts of *Apollo 12* and *Apollo 14* have measured two types of lunar magnetic field: permanent fields due to fossil magnetic material and transient fields due to electric currents generated deep in the interior of the moon. These magnetic measurements yield unique information about the history and present physical state of the moon. The origin of the fossil magnetism is unknown and points to a mysterious epoch in lunar history. The transient fields are induced in the moon by changes in the magnetic field associated with the "solar wind": the stream of electrically charged particles ejected by the sun. The transient fields provide data for computing the electrical resistance and the temperature of the lunar interior. The magnetic measurements demonstrate a promising method for the future scientific exploration and study of bodies in the solar system, such as Mars, that resemble the moon.

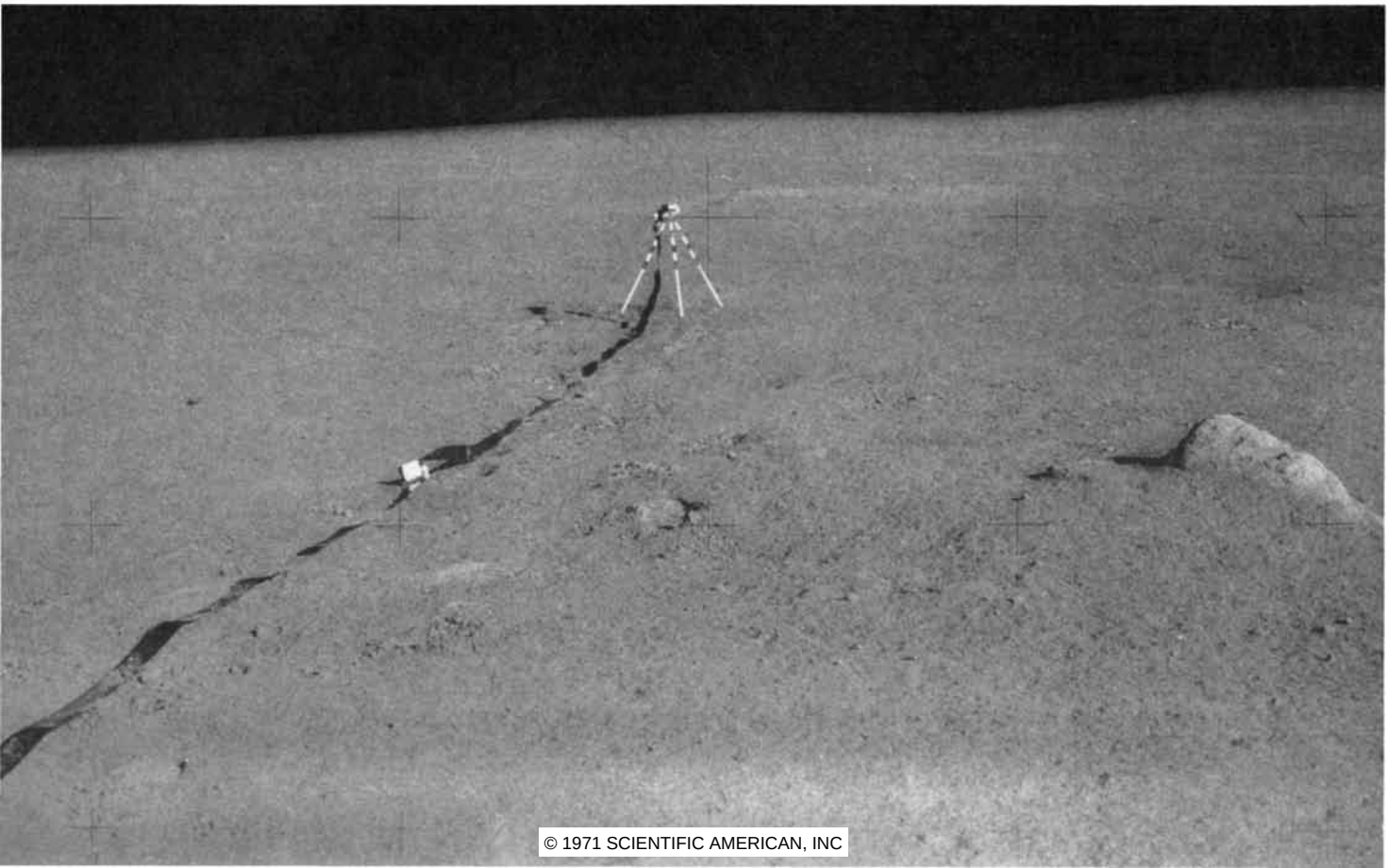
Although the permanent fields measured on the moon so far are less than 1 percent as strong as the magnetic field of the earth, the lunar fields are much stronger than was expected on the basis of magnetometer measurements made earlier by American and Russian spacecraft that either flew past the moon or went into orbit around it. The permanent fields measured by the Apollo instruments vary from place to place on the lunar surface and do not combine to form an overall dipole pattern similar to that of the earth. This indicates that the measured fields are due to local sources. Moreover, the high field strengths indicate that these sources were magnetized by an ambient field much stronger than the one that now exists at the moon. Evidently at some time in the past the moon either possessed a strong magnetizing field or was immersed in one. We are presented with some fascinating questions about lunar history: Did the ancient moon have an earthlike field, generated as the earth's is thought to be by an internal "dynamo"? Was the moon once in an orbit closer to the earth and within the strong terrestrial field? Was the moon magnetized in another part of the solar system and later captured by the earth? The answer to these questions may be recorded in the fossil magnetism of the lunar material. The fact that the moon has changed little for billions of years offers opportunities to investigate magnetic records of the ancient solar system. Similar magnetic information recorded in the earth's crust would have been erased long ago by crustal activity and exchange of magnetic surface material with the earth's molten interior.



LUNAR SURFACE MAGNETOMETER was designed by workers at the Ames Research Center of the National Aeronautics and Space Administration to be emplaced at a single fixed location on the moon, where it can measure three components of the lunar magnetic field continually through the lunar day and night. Three "flux gate" sensors (*color*), each directed at right angles with respect to the other two, are installed at the top ends of the booms about 30 inches above the surface. By remote command from the earth the sensors can be individually flipped 90 degrees or 180 degrees or rotated around their axes by a pulley drive (*not shown*), allowing all three sensors to be aligned simultaneously along any one boom direction. Motors and data-processing equipment are located inside the box, and the entire assembly is encased in a thermal blanket. Heat rejection (during day) and retention (during night) are controlled by an array of parabolic reflectors on two sides of the electronics box. A bubble level and an azimuthal shadowgraph on top of the box enable the astronauts to orient the magnetometer accurately. Power, digital signals and commands are conveyed through a ribbon cable connected to the central-station receiver/transmitter.



APOLLO 14 PORTABLE MAGNETOMETER, a self-contained battery-operated device, was designed to be transported by the astronauts along a surface traverse for the purpose of recording measurements of the permanent lunar magnetic field at different locations. In the photograph above, made by astronaut Edgar D. Mitchell, Jr., the lunar portable magnetometer is shown stowed aboard the two-wheeled mobile equipment transporter. Astronaut Alan B. Shepard, Jr., can be seen working on another piece of equipment behind the transporter. In the photograph below the sensor-tripod portion of the magnetometer is shown deployed about 50 feet from the electronics box, which remains on the transporter during the measurements. After completing all the readings and transmitting them to the earth, the astronaut stows the sensor-tripod assembly, reels up the ribbon cable and continues on the surface traverse.

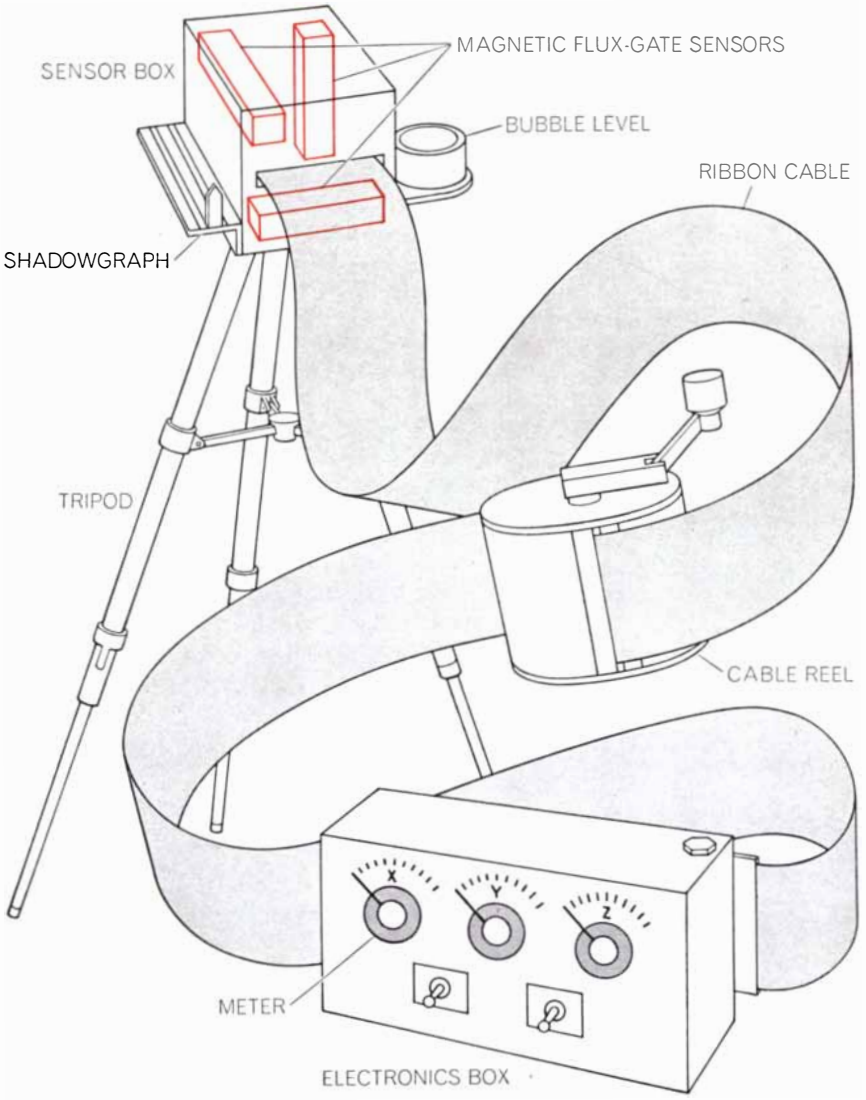


The transient fields generated by electric currents flowing in the lunar interior are associated with the entire moon rather than with any one region. The fields wax and wane rapidly in response to changes in the solar wind. The characteristics of the induced lunar fields depend on the electrical resistance of the moon's interior, and this resistance is related in turn to the temperature of the material. Therefore the magnetometer can be used as an indirect "resistance thermometer" to determine the internal temperature of the moon.

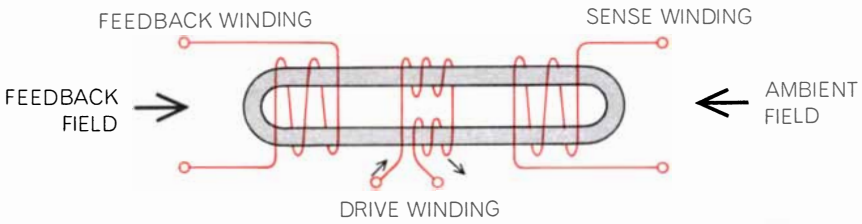
Analysis of the transient fields shows that the temperature of the moon is only about 1,000 degrees Celsius at a depth halfway to the center, or some 3,400 degrees cooler than the earth is halfway to its center. Lunar temperature determinations are consistent with the low seismic activity reported by the seismometers that have been placed on the moon. These experimental results indicate that the moon is not hot and molten inside but is relatively cool, and that its surface has been modified more by the impact of meteorites than by volcanic activity. Additional magnetic-field measurements should yield a more accurate temperature profile of the lunar interior and also shed light on the nature of the internal heat source. Through this knowledge fundamental questions about the origin and evolution of the moon will be answered.

The lunar magnetometer experiment can be regarded as the modern counterpart of the first planetary magnetic experiment conducted by William Gilbert in 1600. Observing that a balanced magnetized needle dips toward the earth at an angle, Gilbert inferred that the earth itself acts as a large magnet. Since then the geometrical and time-varying characteristics of the earth's field have been intensively analyzed to determine the electrical resistance and the temperature of the earth's crust and mantle. Long-term measurements also indicate that the earth's molten interior is rotating at a slightly lower speed than the crust. This differential rotation is thought to produce the dynamo action needed to account for the earth's magnetic field.

With the development of rocket technology it became possible to place magnetometers on earth-orbiting spacecraft, to map the terrestrial magnetic field in nearby space and to determine its interaction with the solar wind. Other magnetometer experiments, conducted by Edward John Smith of the Jet Propulsion Laboratory, were flown on board



LUNAR PORTABLE MAGNETOMETER contains three mutually perpendicular flux-gate sensors, identical with those used in the lunar surface magnetometer, that are housed inside a cubical sensor block and mounted on top of a tripod. The sensor-tripod assembly is connected by means of a 50-foot ribbon cable to the electronics box, which contains a battery pack, electronics equipment and three meters. At each surface site the three magnetic-field vector components are read aloud from the meters by an astronaut and radioed to the earth.



FLUX-GATE SENSOR is used in both types of lunar magnetometer to detect the ambient magnetic field. The sensor consists of an easily magnetized, toroidal Permalloy core that is driven to saturation by an alternating signal at a certain frequency. A sense winding detects the superposition of the drive-winding magnetic field and the ambient magnetic field; as a result a signal of twice the frequency of the driving frequency is generated in the sense winding with a magnitude that is proportional to the strength of the ambient field. The phase of this second harmonic signal with respect to the drive signal indicates the direction of the ambient field with respect to the axis of the sensor. The harmonic signal is electronically processed and sent through the feedback winding to "null out," or cancel, the ambient field within the sensor. Operating at null increases the thermal stability of the sensor.

Mariner spacecraft to the vicinity of Venus and Mars, measuring the interplanetary solar fields en route and determining how these planets interact with the solar wind. No magnetic fields intrinsic to either Venus or Mars were detected.

Magnetic fields associated with the sun and celestial objects outside the solar system have been measured by telescopic observation. As early as 1908 George Ellery Hale observed that the magnetic field of the sun splits lines in the solar spectrum (the Zeeman effect). Subsequently Horace W. Babcock and others used Hale's technique to measure the magnetism of other stars and entire galaxies. Recently James C. Kemp of the University of Oregon has observed an extremely strong field (100 million times stronger than the field of the earth) in a dwarf star in the constellation Draco.

Magnetic studies related to the moon were begun in 1850, when Karl Kreil discovered that the earth's magnetic field varied in a systematic way with the phase of the moon. Later it was shown that the effect was created by the moon's gravitational pull on the earth's atmosphere and ionosphere and had nothing to do with the presence or absence of a lunar magnetic field. It was not until spacecraft able to reach the vicinity of the moon were developed that direct measurements of the moon's field finally became possible.

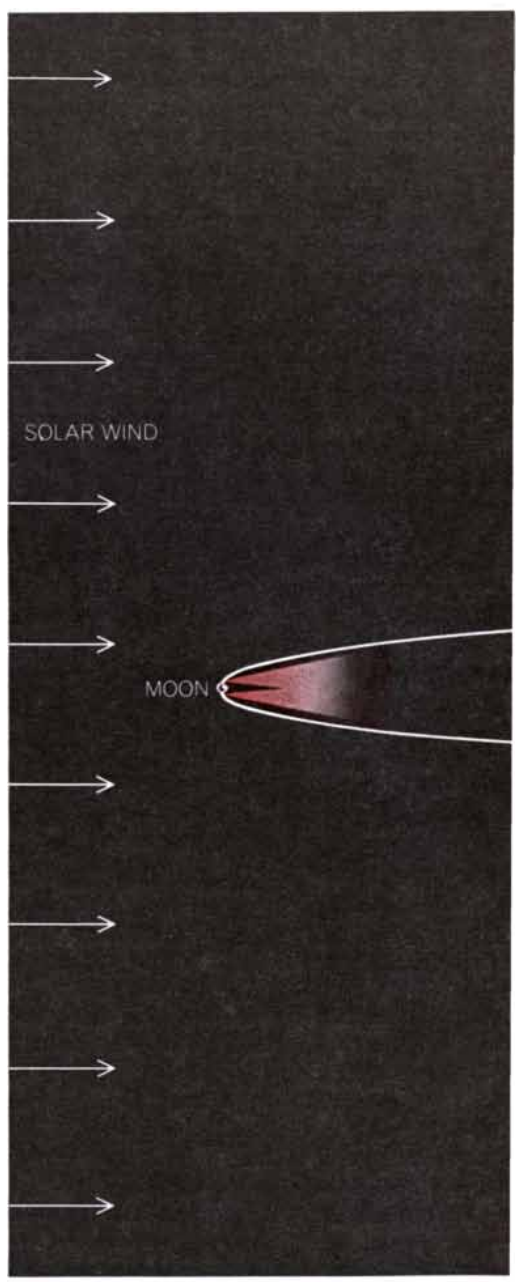
Since no rigorous theory has evolved that satisfactorily explains the earth's permanent magnetic field, it is not surprising that no one predicted the magnitude of the lunar permanent field. Investigations of other lunar properties, such as optical measurements of changes in the moon's surface temperature and studies of the moon's shape and dynamic behavior, have indicated, however, that the moon does not have an internal dynamo.

Theoretical predictions of other lunar magnetic properties were made before the manned lunar missions by several investigators, including F. Curtis Michel of Rice University, Thomas Gold of Cornell University and John R. Spreiter of Stanford University. They attempted to predict how the moon would interact with high-speed particles in the solar wind if, at one extreme, the moon was completely nonmagnetic and if, at the other, it had a magnetic field as strong as the earth's. (The earth's field at the Equator is about 30,000 gammas.) At the nonmagnetic extreme the moon would act simply as an opaque sphere, intercepting only the charged solar particles that happened to strike it and allowing the others to pass undeflected [*see illustration at right*]. If the moon had a significant overall field, however, the particles of the solar wind would be deflected around the lunar sphere in a complex manner. Moreover, a shock front would be created on the upwind side of the moon, in rough analogy to the shock front created ahead of a supersonic aircraft.

Several possibilities lying between these extremes were considered. If the moon did not have a global field of at least 40 gammas (as was later found to be the case), time-varying magnetic fields induced inside the moon could still substantially perturb the flow of the solar wind. Two basic types of induced lunar field are possible. The first is a type of field that would be induced in the moon by the motion of the solar-wind field past the moon. This process was mentioned by Gold and was developed theoretically by David S. Colburn and Charles P. Sonett of the Ames Research Center of the National Aeronautics and Space Administration. The second possibility is a time-varying dipolar field, which would be induced in the moon by temporal changes in the solar-wind field [*see illustration on page 71*]. In this process eddy currents are set up inside the moon by time-varying solar-wind fields in much the same way that currents are generated in the secondary coil of a transformer by changing fields in the primary coil [*see illustration on page 72*]. This latter mechanism was suggested by one of us (Dyal) in 1966 and was developed theoretically by Joel Blank and William Sill of Bellcomm, Inc., in 1969. Magnetometer measurements made on the surface of the moon have subsequently proved that the dipolar eddy-current field is indeed the dominant lunar induced field.

The measurement of magnetic fields in the vicinity of the moon began in January, 1959, when the Russian spacecraft *Luna 1* carried a magnetometer to within several hundred miles of the moon. In September, 1959, *Luna 2,* also equipped with a magnetometer, crashed into the moon. The instrument aboard *Luna 2* set an upper limit of 100 gammas for a possible lunar field at an altitude of 30 miles above the moon's surface. In April, 1966, *Luna 10,* carrying a magnetometer 10 times more sensitive than *Luna 2*'s, was successfully placed in a lunar orbit that came to within 220 miles of the moon. The *Luna 10* magnetometer measured a magnetic field varying between 24 and 40 gammas in the neighborhood of the moon. This field, which was correlated with changes in magnetic activity at the earth's surface, was interpreted by the Russian investigator L. N. Zhuzgov as indicating the existence of a weak lunar magnetosphere.

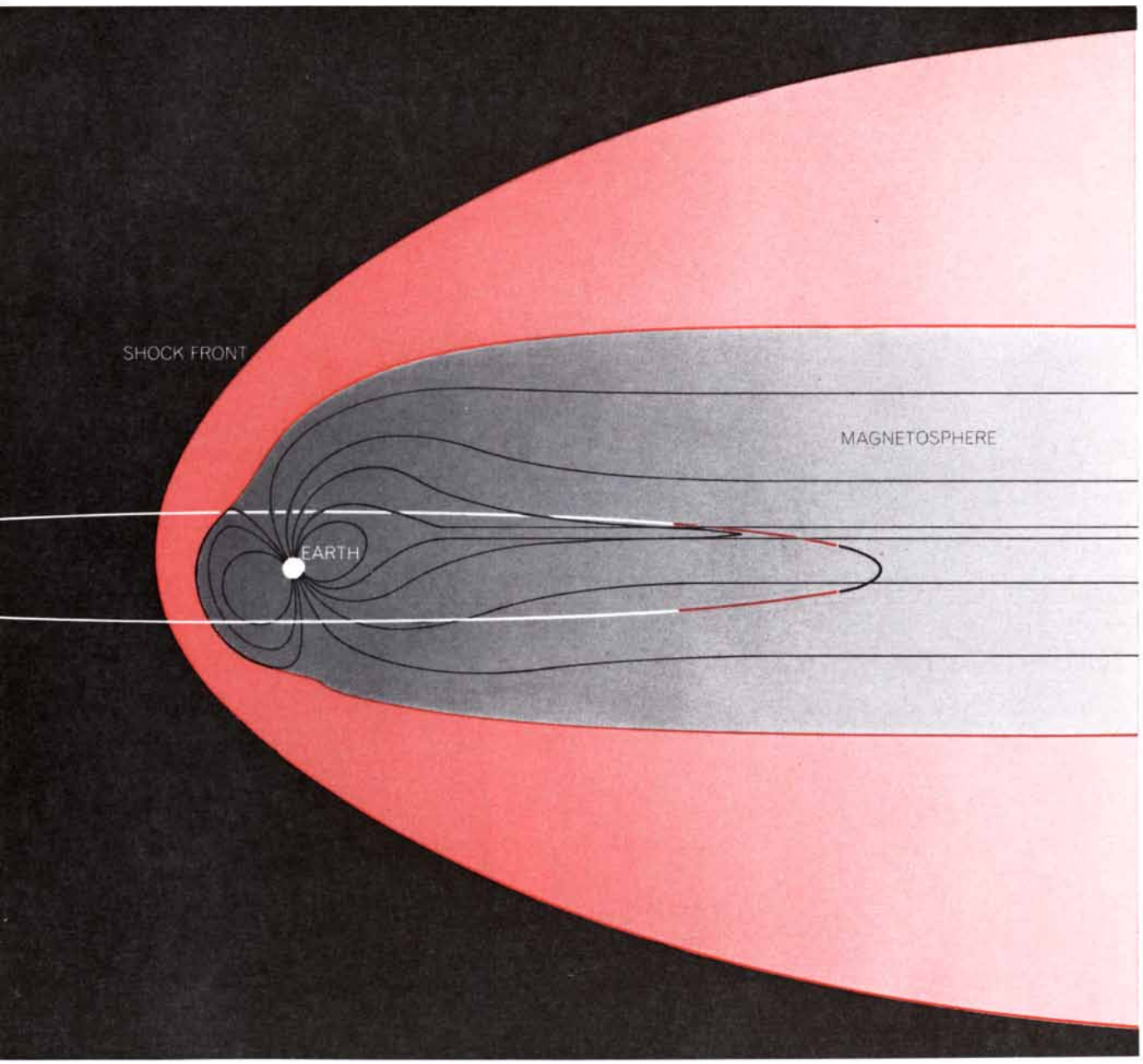
A year later (July, 1967) the U.S. placed *Explorer 35* in orbit around the



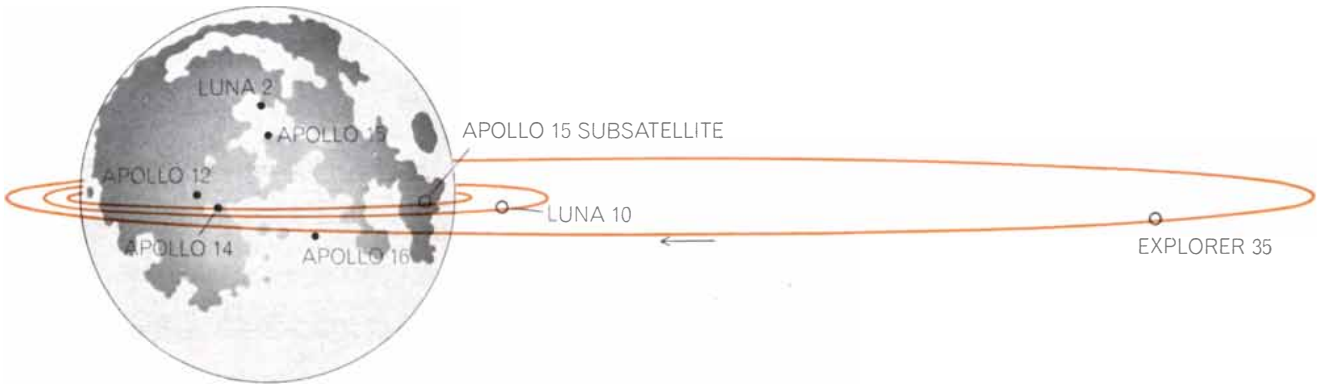
MAGNETIC ENVIRONMENT of the moon varies considerably depending on the moon's orbital position. The "solar wind," made up of particles blown outward from the sun at supersonic speeds, sweeps the

moon with two magnetometers aboard. The spacecraft is still operating in a trajectory that takes it to within 520 miles of the moon's surface. With instruments whose sensitivity is ±.2 gamma *Explorer 35* successfully measured magnetic properties of the solar-wind cavity downstream from the moon. *Explorer 35* failed, however, to detect the lunar magnetosphere indicated by *Luna 10* measurements or the shock front and induced-field configuration suggested by Gold. In an analysis of the *Explorer 35* results Sonett concluded that if a permanent lunar field existed at all, its magnitude would be less than two gammas, a disappointingly low value, at an altitude of 520 miles.

This was the situation, with all signs pointing to a magnetically dead moon, when we finished testing the magnetometer package scheduled to be carried to the Ocean of Storms by *Apollo 12* in November, 1969. One can imagine our surprise and delight when the instrument measured a permanent magnetic field of 38 gammas at the *Apollo 12* landing site. Then in February, 1970, Alan B. Shepard, Jr., and Edgar D. Mitchell, Jr., of *Apollo 14*, working with a portable magnetometer, measured magnetic fields of 43 gammas and 103 gammas at two different locations near their landing site, some 110 miles east of where *Apollo 12* had landed.



earth's magnetic field into a tubular shape known as the magnetosphere (*gray area*). The magnetosphere in turn acts as a barrier to the solar wind, causing a bow-shaped shock front to form ahead of the earth. When the moon is immersed in the magnetosphere, the moon's magnetic environment is dominated by the comparatively steady magnetic field of the earth. In the free-streaming solar wind the moon is subjected not only to the solar magnetic field but also to plasma waves traveling from the sun. The intermediate magnetosheath region (*light-colored area*) is characterized by erratic solar-particle flow and the most turbulent fields of the lunar orbit.



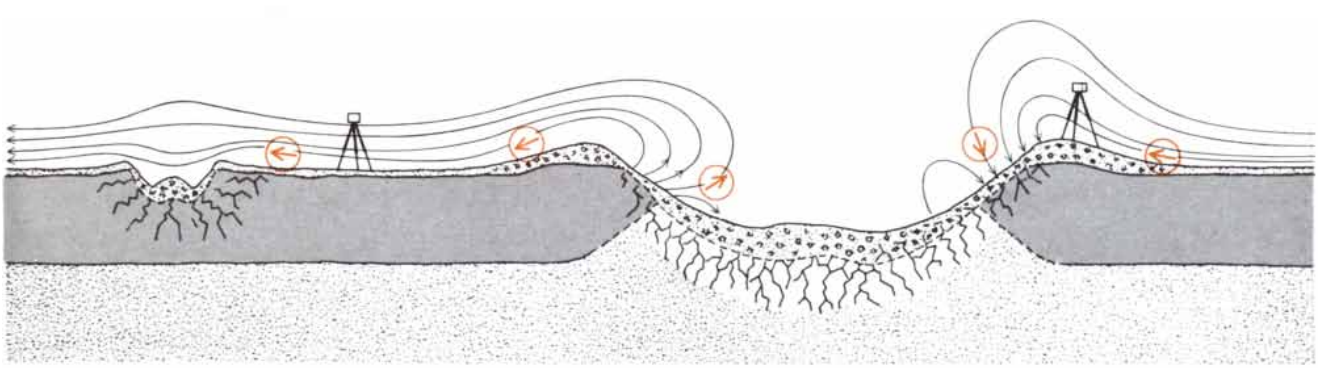
LOCATIONS of past lunar magnetometer experiments and those scheduled by the U.S. through 1972 are indicated on this map of the front half of the moon. The Russian *Luna 2* spacecraft, which crashed into the moon in 1961, and the *Luna 10* orbiter, launched in 1966, made the earliest magnetic measurements of the lunar environment but were not sensitive enough to detect either the steady or the transient lunar fields. The U.S. *Explorer 35* orbiter, placed in lunar orbit in 1967, has made possible detailed study of the effects of the moon on the solar wind, but it is too distant from the moon to measure the lunar fields. The magnetometers placed on the moon at the *Apollo 12* and *Apollo 14* sites respectively in 1969 and 1971 have found the moon to be much more active magnetically than was anticipated. Magnetometers scheduled for future Apollo missions to the moon should provide more detailed information.

The magnetometers placed on the moon by the astronauts of *Apollo 12* and *Apollo 14* were designed at the Ames Research Center. Sonett originated the *Apollo 12* experiment and one of us (Dyal) extended this concept to a network of stationary and portable magnetometers for a series of subsequent Apollo missions. The instrument used in the *Apollo 12* experiment was designed and fabricated by the Philco-Ford Corporation under the direction of John Keeler of the Ames Research Center. It was integrated into an array with four other experiments, collectively called the Apollo Lunar Surface Experiments Package (ALSEP), by the Bendix Corporation. The *Apollo 14* portable magnetometer was fabricated by an Ames Research Center team led by Carle Privette and was integrated into the Apollo spacecraft by the Manned Spacecraft Center of NASA.

The design of the *Apollo 12* instrument took into account the magnetometer readings provided by *Explorer 35* and *Luna 2*. In the Apollo instrument three components of the vector magnetic field are detected by three "flux gate" sensors located at the ends of three mutually perpendicular booms [*see illustration on page 63*]. The sensors are separated so that the uniformity of the local magnetic field can be determined by measuring field values at each sensor location. Each sensor weighs less than an ounce and operates on 15 milliwatts of power.

A flux-gate sensor consists of an easily magnetized core (Permalloy) with three windings: one to drive the core to saturation in the presence of the ambient magnetic field, another to sense the sum of the driving field and the ambient field, and a third to cancel out the ambient field inside the sensor. A signal with twice the frequency of the driving frequency is induced in the sense winding; the magnitude of this signal is proportional to the ambient field [*see bottom illustration on page 65*]. Each sensor has a frequency response ranging from zero to three hertz (cycles per second) and an angular response that is proportional to the cosine of the angle between the magnetic-field vector and the sensor axis. The instrument can be set at ranges of ±100 gammas, ±200 gammas or ±400 gammas, with an accuracy of .2 percent in each range.

The electronics subsystem is self-contained except for power supplied by a radioactive-isotope thermoelectric generator and timing and telemetry supplied by the ALSEP central station. The subsystem weighs six pounds, occupies a volume of 300 cubic inches and has some 6,400 electronic parts. The electronic circuits drive the flux-gate sen-

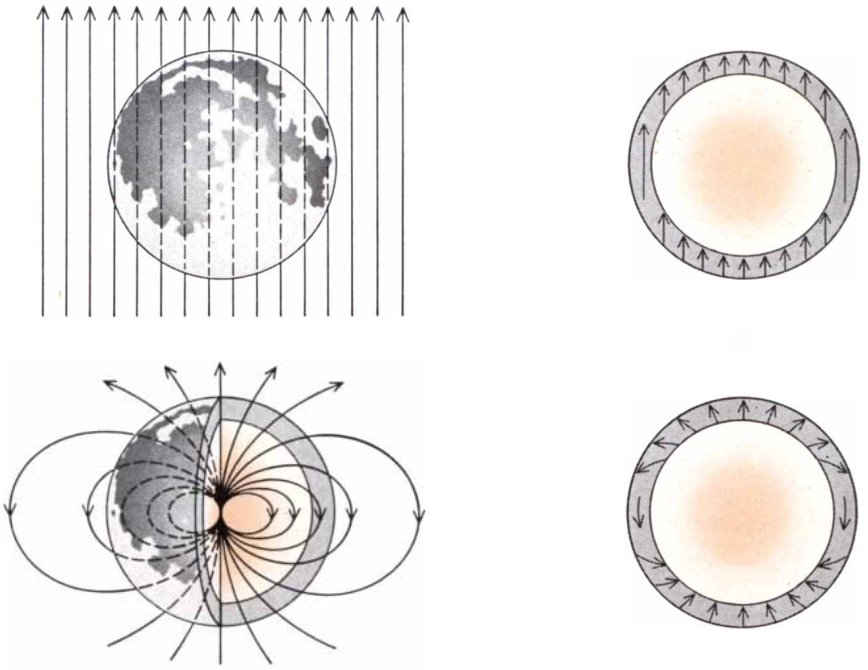


NONUNIFORM STEADY MAGNETIC FIELD was measured by the *Apollo 14* astronauts at two locations separated by about three-quarters of a mile, indicating that strongly magnetized material exists in that vicinity. The nonuniformity could be explained by the existence of a subsurface slab of material (*dark gray area*) that was uniformly magnetized at one time but has subsequently been altered by local processes such as shock demagnetization from meteorite impacts. The moon has little or no global dipole field; it appears rather that the permanent lunar magnetic field is dominated by local features such as the one illustrated here and that it varies from place to place. Therefore a compass (*colored arrows*) would probably not be useful for navigation on the moon.

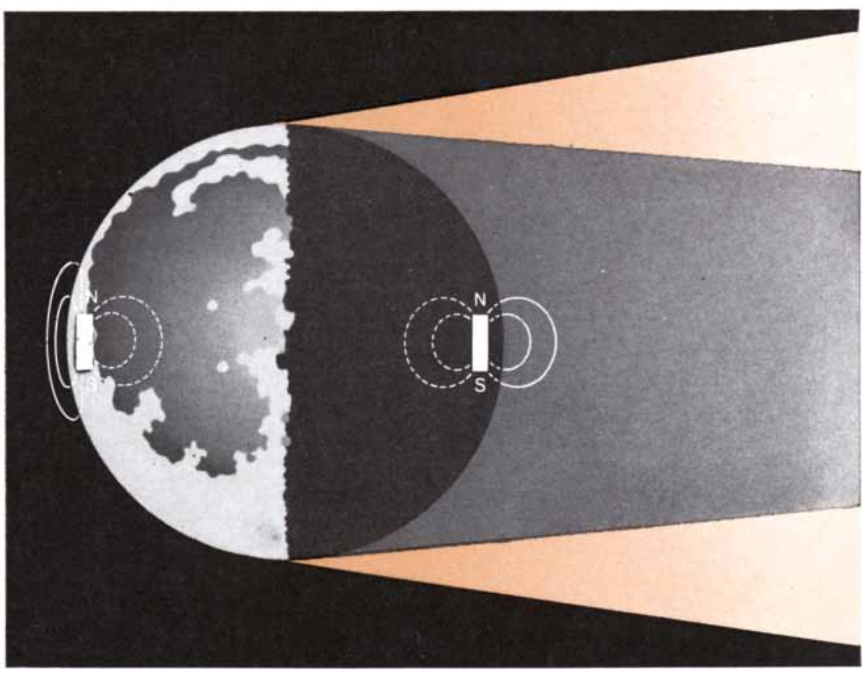
sors, measure the orientation of the instrument, drive three motors for sensor orientation and process the field data before radio transmission to the earth. The instrument is wrapped with thermal insulation and has directional heat reflectors to radiate heat out into space during the lunar day; resistance heaters supply heat during the lunar night. In this way the temperature of the instrument is maintained between −25 degrees C. and 75 degrees C. even though the temperature of the lunar surface on which it sits varies from about −150 degrees to 120 degrees.

During the *Apollo 12* mission the magnetometer was placed in operating position by astronaut Alan L. Bean at 23.35 degrees lunar west longitude and 2.97 degrees south latitude. Bean leveled the instrument with a bubble level and aligned it in azimuth by adjusting a shadowgraph to read within marked preset values. The instrument was then activated by radio command from the Manned Spacecraft Center at Houston, and measurements transmitted back to the earth showed that the instrument was operating successfully.

The success of the *Apollo 12* magnetometer experiment lead to the design of a portable magnetometer to measure the local permanent fields at the *Apollo 14* landing site. The *Apollo 14* magnetometer was designed to be entirely self-contained, so that the astronauts could measure the steady magnetic field at different points along their traverse. This design makes it possible to measure the changes in magnetic field over distances on the order of a mile rather than, as in the case of *Apollo 12*, 4.5 feet. Three mutually perpendicular flux-gate sensors, identical with those in the *Apollo 12* instrument, are mounted on top of a tripod; the sensor-tripod assembly is connected by a 50-foot cable to an electronics box that contains a battery pack, the electronics and three meters [*see top illustration on page 65*]. The astronauts read the meters and report the needle positions back to the earth by radio. The 50-foot separation minimizes the effect of the magnetic fields associated with the electronics and the astronauts' backpack. The instrument weighs 11 pounds, operates on 1.5 watts of power and functions over the temperature range from zero degrees C. to 50 degrees. It is designed to filter out high-frequency fluctuations in the solar-wind field. Two magnetic-field measurements were made by Shepard and Mitchell during their second exploratory excursion on the moon. The first measurement was obtained 350 yards to the



MAGNETIC MEASUREMENTS SUGGEST that at one time in the lunar past ambient fields much stronger than those currently observed existed over much or all of the lunar surface. One explanation is that the moon may have taken a "magnetic snapshot" of an early stage in the evolution of the solar system. If the crust of the entire moon was magnetized at that time, different global patterns of magnetized material should exist depending on whether the ancient ambient field originated outside or inside the moon. An external field (*top left*) that was stronger than the solar or terrestrial fields that exist at present would tend to magnetize the crustal material unidirectionally (*top right*), whereas an internal field (*bottom left*) would result in a varied pattern of magnetization (*bottom right*).



SOLAR WIND COULD COMPRESS a steady lunar magnetic field, forming a bow-shaped shock front, provided that the steady field is strong enough and its source is large enough. A global lunar dipole field, if it exists at all, would be too weak to create a shock front like the shock front of the earth. Instead the solar wind impinges directly on the lunar surface on the daytime side and forms a turbulent "wake" downstream (*light-colored area*), which bounds a "cavity" essentially empty of solar-wind particles (*gray area*). A localized steady field of sufficient strength and extent, although unaffected while in the cavity, would be compressed while on the daytime side of the moon. Such a steady-field compression has been observed at the *Apollo 12* site for data gathered during a solar magnetic storm.

east of the lunar module; the second, .7 mile farther east near the rim of a crater.

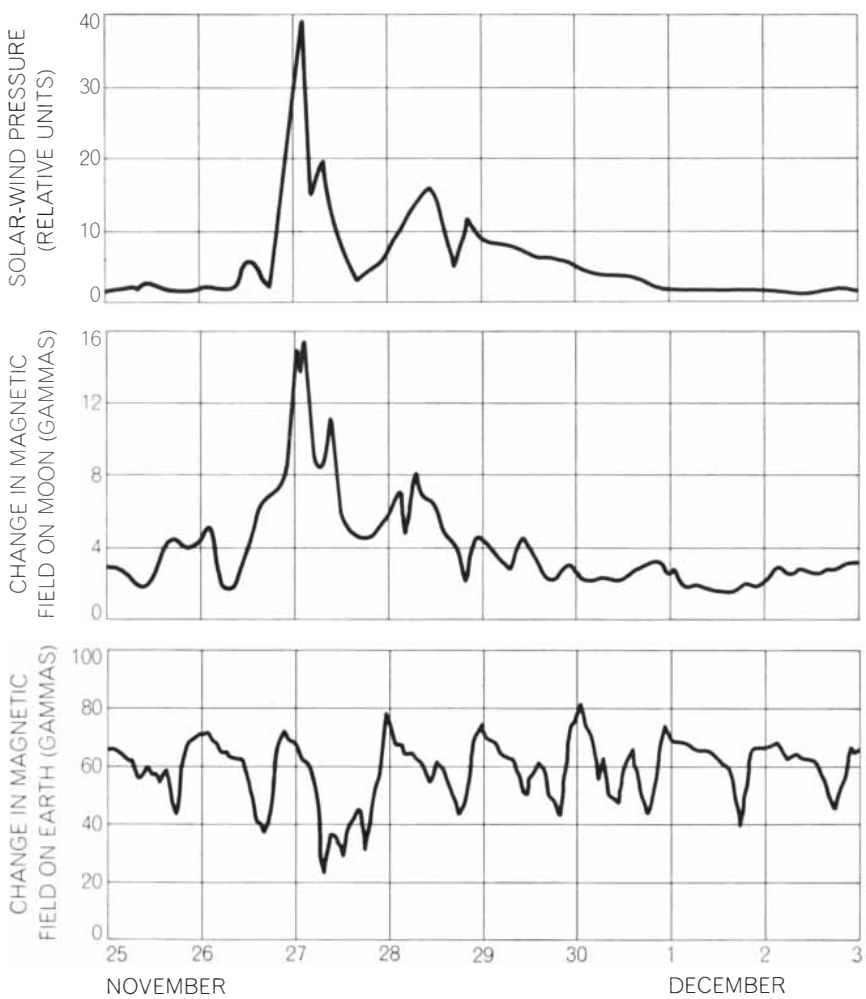
Measurements returned from the Apollo magnetometers on the lunar surface have yielded much more information than was anticipated. Not only were steady magnetic fields measured that were up to 25 times stronger than had been predicted but also well-defined lunar induction fields were observed that enabled us to calculate values for the electrical resistance and temperature of the moon. The variation in readings from less than 40 gammas to more than 100 gammas told us that our instruments were measuring fields due to highly magnetized local sources rather than an overall lunar dipole field.

The steady-field value of approximately 38 gammas—some 10 times the maximum predicted value—radioed back to the earth just 40 minutes after Bean had unfolded the *Apollo 12* magnetometer was so unexpected that we hesitated to believe it until we had calibrated the instrument by sending radio commands from Houston back to the experiment package. Three days after the magnetometer was turned on, the moon passed into the tail of the earth's magnetosphere, where the solar wind is excluded by the earth's magnetic field. During the time when the moon was inside this region of very steady magnetic fields we sent commands to the instrument that enabled it to function as a gradiometer. In this mode the sensors were rotated by motors so that all three vector-field components could be measured sequentially at each of the three sensor locations. The three measurements showed that the field varies less than the instrument's maximum sensitivity of .2 gamma over the 4.5-foot distance between any two sensors.

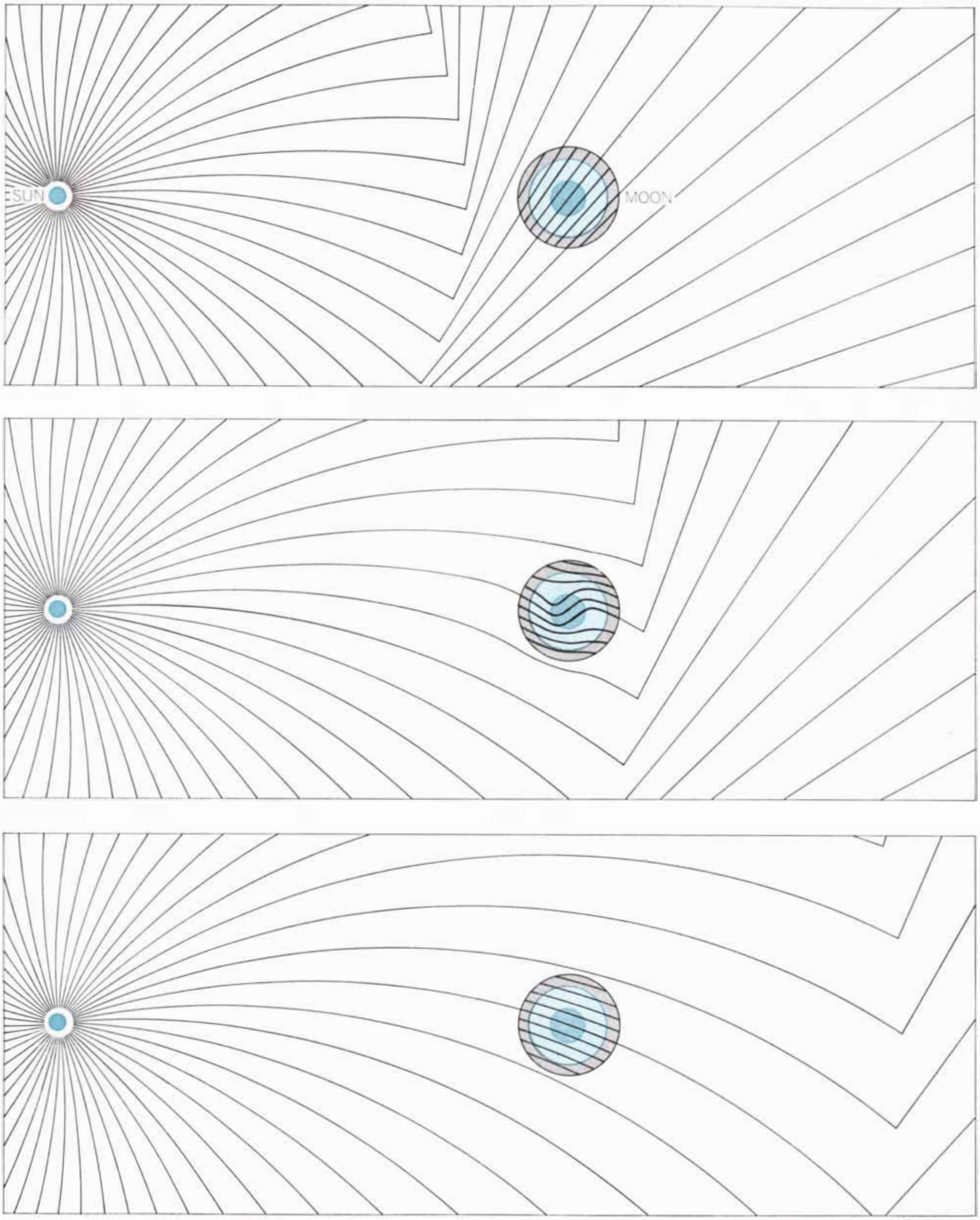
The spatial uniformity of the field indicates that if the highly magnetized source were a single meteorite buried near the surface, it would have to be more than 200 yards away from the instrument. On the other hand, the upper limit placed on the permanent field by *Explorer 35* required that the source be within 125 miles of the *Apollo 12* instrument. One can compute the minimum and maximum dimensions of the source if it is assumed to be a sphere with the same remanent magnetization as that of lunar samples brought back to the earth. We calculated that if the hypothetical source lay 200 yards from the magnetometer, it would be 50 yards in diameter, and if it were 125 miles away, it would be 30 miles in diameter.

Another observation that sheds light on the dimensions of the steady field is the compression of the field by the solar wind. Aaron Barnes and Patrick Cassen of the Ames Research Center predicted that the 38-gamma permanent magnetic field would be compressed if its effective extent above the surface were more than about six miles. Such a compression of the 38-gamma field at the *Apollo 12* site has been observed during times when the moon is showered by a high density of solar-wind particles [*see bottom illustration on preceding page*]. The solar-wind density was measured by Conway Snyder and Douglas Clay of the Jet Propulsion Laboratory.

The magnetic-field measurements made at the *Apollo 12* and *Apollo 14* sites, together with the high magnetic remanence found in rock samples returned from all the Apollo sites (including the *Apollo 11* site some 900 miles to the east of *Apollo 12*), demonstrate that the moon has been magnetized in widely dispersed regions. John Mihalov of the Ames Research Center has reexamined the magnetometer and charged-particle data from *Explorer 35* and has concluded that several magnetized areas may exist on both the near and the far side of the moon. Thus we have strong evidence that much of the lunar surface—perhaps even a crustal shell around the entire moon—was magnetized at some time in the past. Evidently ambient magnetic fields much stronger than those observed today existed over much or all of the lunar surface. It may be that



EFFECTS OF SOLAR STORM on the lunar magnetic field at the *Apollo 12* site and on the terrestrial magnetic field at a site on the surface of the earth are contrasted in this illustration. The top curve shows the rise in solar-wind pressure that accompanies the arrival of solar-storm particles in the vicinity of the earth-moon system. The middle curve shows that the horizontal component of the *Apollo 12* steady field rises in direct proportion to the solar-wind pressure. The bottom curve shows that the terrestrial field intensity, in contrast, exhibits a corresponding decrease during the solar storm. This decrease in the earth's total field is believed to be caused by the formation of earth-encircling ring currents involving charged particles trapped in the earth's magnetosphere; no known mechanism exists for the formation of an analogous ring current around the moon.



TRANSIENT MAGNETIC RESPONSE of a hypothetical three-layered model of the moon is represented for a case in which a directional discontinuity in the solar magnetic field travels outward past the moon at a supersonic velocity. The three layers are characterized by successively decreasing electrical resistance with distance into the moon. The top drawing shows the moon immersed in a steady solar field before the arrival of the discontinuity. The middle drawing shows the magnetic field through the moon immediately after the passage of the discontinuity. Eddy currents have been induced in the moon according to Lenz's law, which states that in an electrically conducting medium, currents and fields will be induced that tend to oppose any change in the original ambient magnetic field. Such eddy currents persist longest in regions of low resistance, so that the original magnetic-field orientation is maintained longest in the core. Eventually all the eddy currents decay and distortions of the total field disappear (*bottom drawing*).

the moon has taken a magnetic snapshot of an early evolutionary phase of the solar system.

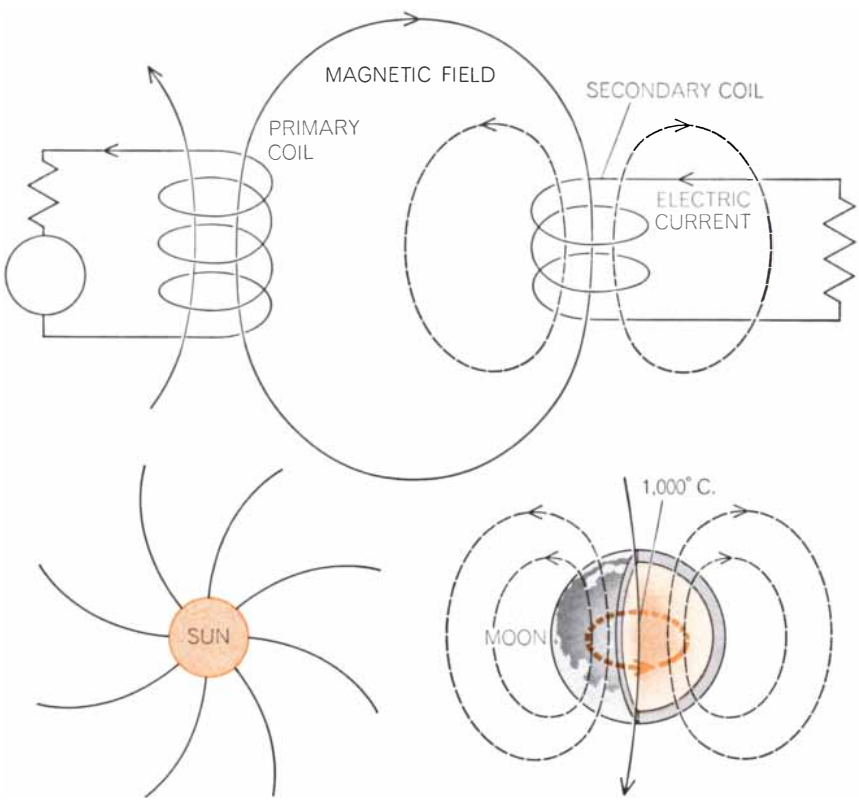
The magnetic-field measurements made at the *Apollo 12* and *Apollo 14* sites are strikingly similar: all the vectors point down and toward the south, and their magnitudes correspond to within a factor of three. This suggests that the two *Apollo 14* sites and possibly the *Apollo 12* site are located over a slab of material that was uniformly magnetized at one time. Subsequently the magnetism of the slab could have been altered by local processes, for example by tectonic activity, by fracturing or by demagnetizing shock of meteorite impacts [*see bottom illustration on page 68*].

The material near the lunar surface was probably magnetized at the time of the moon's crustal solidification some 3.7 billion years ago. The remanent magnetization in the samples from *Apollo 11* and *Apollo 12* would have required an external magnetizing field stronger than 1,000 gammas; ambient fields of this magnitude have not been measured in space near the moon. The source of the ancient ambient field could have been external to the moon (produced by the sun or the earth) or inside the moon (produced by dynamo action or internal electric currents).

The earth's magnetic field could have magnetized the lunar material if the earth's field was much stronger in the past or if the moon's orbit was once much closer to the earth. If the terrestrial field was never stronger than it is today, the moon would have had to approach to within two or three earth radii in order for the moon to be subjected to a magnetic field of 1,000 gammas; this would be close to the "Roche limit," where tidal forces would break up the moon. For the required magnetizing field to have been produced by an intrinsic lunar dynamo, the moon would have had to possess both a hot core and a fairly high rate of spin at the time the surface material cooled below its Curie temperature (the temperature where magnetization is "frozen" into the material). This hypothesis requires some mechanisms, unknown at present, for lowering the temperature of the lunar interior and slowing the spin rate to their present values. Our knowledge of the moon is still too limited to allow a choice among these and other hypotheses. One hopes that further mapping of the moon's steady fields during future surface and orbital missions will elucidate the moon's "magnetic epoch" and solve one of the more interesting puzzles of lunar history.

Whatever models may be devised to explain this epoch, they will have to take account of the moon's low magnetic permeability. There has been speculation that portions of the lunar interior might consist of a high percentage of iron, in which case the moon could exhibit substantial permeability. Our analysis of magnetometer data from *Explorer 35* and *Apollo 12* shows otherwise. If the moon were significantly permeable, it should measurably distort the lines of magnetic force in the lunar environment. Little or no distortion is observed; the moon's relative permeability appears to be about as low as that of free space. We conclude that the moon does not as a whole possess the properties of a large magnet.

Those of us engaged in developing the magnetometer experiments had hoped that measurements made at the *Apollo 12* site might provide clues to the electrical conductivity of the moon's interior from which we might be able to infer the moon's internal temperature. We could not predict in advance, however, whether readings obtained during the lunar day or the lunar night would be most useful for the purpose. It turned out that the most easily interpreted measurements were those made during the lunar night. During the day the solar wind compresses the 38-gamma permanent field, and this compression is proportional to the highly variable density of the solar wind. Thus if the daytime data are to be accurately analyzed, simultaneous solar-wind effects must be taken into account. During the night the moon can be treated as an electrically conducting sphere in a vacuum.



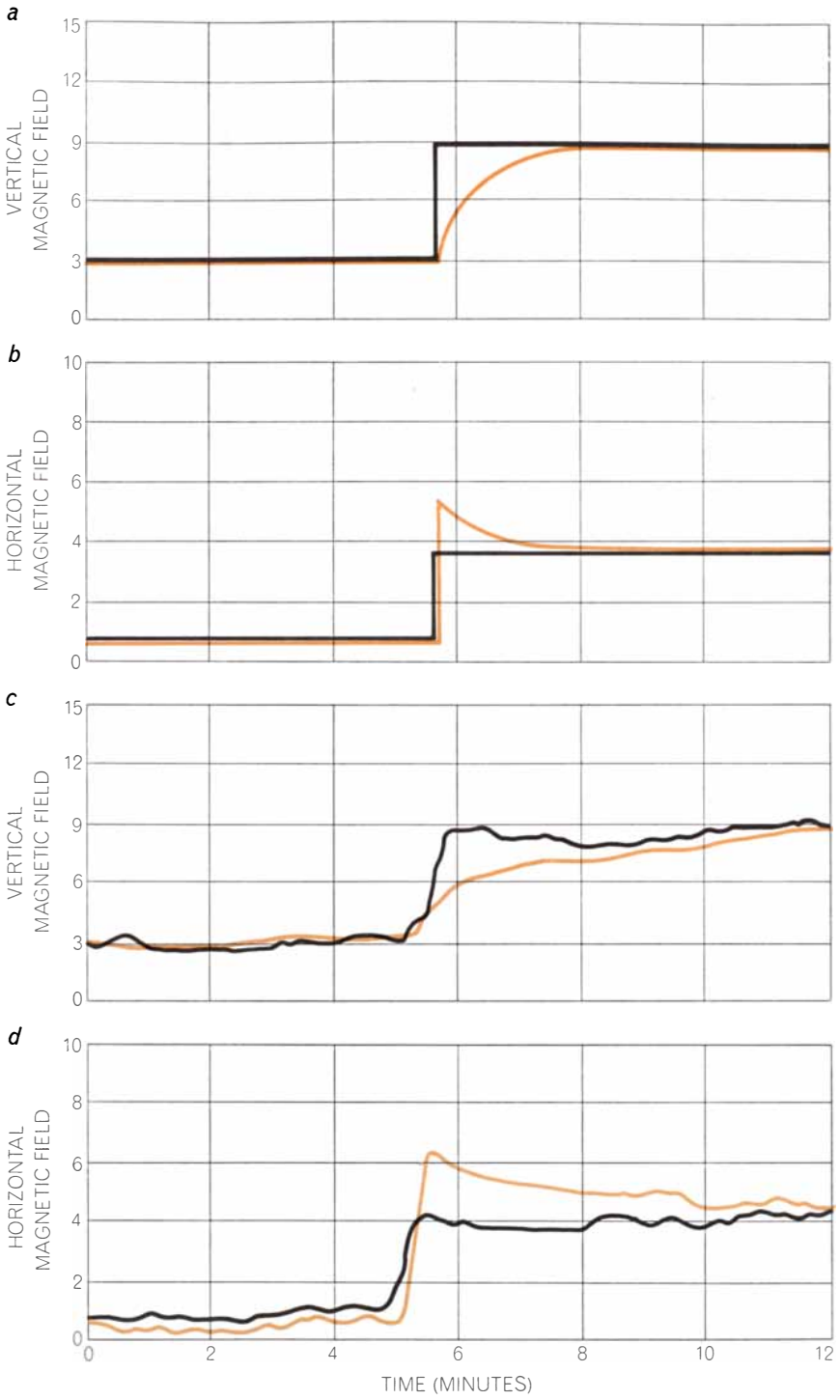
EDDY CURRENTS AND MAGNETIC FIELDS can be induced in the moon in a manner analogous to the way currents and fields are set up in a passive resistive-inductive secondary of a transformer (*top*). The primary of the transformer has a power source that can drive a current in the coil and change its magnetic field. Field lines from the primary thread the secondary coil, and any change in the primary field results in the formation of a dipole field in the secondary. The time necessary for the decay of the secondary current and secondary field depends on the inductance and resistance of the secondary circuit. By analogy the sun acts as a primary in the sense that it is a source of a magnetic field and has the capacity to change that field. A sudden change in the solar field accordingly results in the induction of a corresponding current and a dipole field in the moon (*bottom*). The decay time of the lunar current and field is a function of both the electrical resistance and the size of the moon. The temperature of the moon's interior can be estimated from calculations of the electrical resistance for various assumed lunar compositions. For example, on the assumption that the moon is composed entirely of pure olivine, the temperature of the moon increases with depth to about 1,000 degrees Celsius in the central region.

Because the sun and the moon are electromagnetically coupled by the solar wind, the sun acts as the primary coil of a transformer and the moon acts as a secondary coil. Any sudden change in the strength of the primary magnetic field causes a change in the electric current and magnetic field induced in the secondary. The time required for the secondary current and field to decay is a function of the electrical resistance and size of the secondary, that is, of the moon itself.

The magnetometer on the surface of the moon at night and the magnetometer aboard *Explorer 35* give us simultaneous readings that can be compared whenever there is a "step" (sudden) transient in the strength of the magnetic field carried by the solar wind [*see illustration at right*]. In general the surface fields that point vertically out of the moon respond slowly to rapid changes in the field external to the moon, whereas fields that point along the surface respond rapidly and are amplified compared with simultaneous values recorded in nearby space.

We have examined the magnetometer readings produced by more than 100 step transients in the solar wind. The analysis clearly shows that the entire moon responds to each transient and that the induced currents start immediately near the surface and diffuse slowly into the deep interior. Our calculations show that the outer portion (about 40 percent of the distance to the center) of the moon has a resistance of approximately 10,000 ohm-meters and that the inner region, or core, has a resistance of only 100 ohm-meters. Comparable resistance values for the earth are respectively .1 ohm-meter and .00001 ohm-meter.

The calculated resistance values of the moon can be used to estimate the temperature at various depths in the moon. For this one uses laboratory measurements of resistance as a function of temperature (and pressure) for the most likely lunar constituents. Because such measurements are made with pure minerals they can at best only approximate the moon's actual composition. If one assumes that the chief constituent of the moon is peridotite (a common mineral in the mantle of the earth), the observed electrical resistance of the moon corresponds to a maximum temperature of between 600 degrees and 1,000 degrees C. for the bulk of the moon's interior. These temperatures are only approximate, but they support the hypothesis that for most of its history the moon has not had a large molten core.



MAGNETOMETER DATA, recorded as a sharp solar-field directional discontinuity passed by the moon, were used to calculate the moon's internal electrical resistance. The data were gathered while the surface magnetometer was on the dark side of the moon, so that the compressive effects of the solar wind did not have to be considered. Theoretical solutions for dark-side surface fields on a moon of homogeneous internal resistance are shown in top two graphs. The black curves show the "step" solar-field change in two external magnetic-field components (vertical and horizontal to the lunar surface). The colored curves show the theoretical total surface-field values characteristic of field components that are vertical (*a*) and horizontal (*b*). The bottom two graphs show actual magnetometer data samples. The black curves represent measurements of a solar-field change recorded by the *Explorer 35* lunar orbiter. The colored curves are simultaneous vertical (*c*) and horizontal (*d*) components of lunar surface fields measured by the *Apollo 12* magnetometer. The deviations of the data from the theory are believed to be caused by a core of lower resistance than that of the outer layers of the moon. (The deep resistance can be calculated from the magnitude of the deviation.) The distinct difference between vertical and horizontal components illustrates that the moon is responding as an entire body to changes in the solar magnetic field.

FLIGHT ORIENTATION IN LOCUSTS

The locust is equipped with a sturdy "flight motor" that responds swiftly to start and stop signals from sense organs. Other elegant systems keep the insect from rolling, pitching and yawing in flight

by Jeffrey M. Camhi

Shine a light on a scavenging nocturnal cockroach and it will scurry away. Leave a light bulb burning on a warm summer night and a swarm of insects will soon appear. The capacity of insects and other animals to detect, localize and respond to environmental stimuli has long presented a puzzle to biologists. Nineteenth-century explanations of animal behavior were often colored by anthropomorphism. Insects were said to approach lights because of "curiosity" or to evade them because of "innate fear." By the early part of the 20th century most biologists had abandoned explanations based on the untestable "emotions" of animals. In reaction to the anthropomorphic views, and extrapolating freely from experiments conducted by physiologists such as Sir Charles Sherrington, a number of biologists then adopted the radical hypothesis that animal behavior was almost wholly reflexive. On this view the continually changing array of stimuli that an animal encounters as it moves through its environment was thought to produce a large part of the animal's repertory of behavior by reflexes and their mutual interactions. The central nervous system, particularly the spinal cord and the lower brain, was viewed as being little more than a passive recipient and integrator of multiple sensory inputs.

Recent investigations have made the reflex hypothesis untenable. Nerve networks located in the central nervous system of both vertebrate and invertebrate animals now appear capable of numerous and varied functions that are essentially independent of sensory controls. Although this independent mode of operation has now been demonstrated in many species, it shows up with special clarity in the nerve networks that mediate flying in the locust. My own work at Cornell University, and earlier at Harvard University, has been concerned with the controls that enable a locust to orient itself in flight. Like an airplane, an insect can roll around its longitudinal axis, pitch around a horizontal axis perpendicular to its direction of flight or yaw around a vertical axis [*see illustration at top left on page 78*]. The insect I have used in my studies is the hardy desert locust (*Schistocerca gregaria*), whose annual migrations in Africa and the Middle East cause widespread destruction of crops. My work shows that central-nervous-system networks, independent of the senses, can function in a wide variety of ways and give rise to kinds of behavior difficult to explain on the hypothesis that the nervous system is largely a reflexive mechanism.

The trend from anthropomorphism to reflexive interpretations of animal behavior began late in the 19th century, largely with the work of the French biologist Jacques Loeb. He sought to trace animal movements to the same physical and chemical forces that his mentor, the German botanist Julius von Sachs, had used to explain tropisms: plant movements in response to forces such as gravity. Loeb ultimately described animal orientation in terms of the "tonus theory," which attributes the functional symmetry of animals to their bilateral structural symmetry. On this theory the chemical and mechanical forces acting on the nerves and muscles of an animal's two sides are ordinarily in balance; if the balance is upset by external stimuli, the animal responds in such a way as to restore it.

For example, to explain how a walking insect might automatically and without "fear" turn away from a light source and escape, one simply had to postulate that the amount of light reaching the insect's eyes is unequal. If the extra light happens to fall on the right eye, it generates an excess of nerve impulses in the pathways leading to the legs on the insect's right side, thereby causing the animal to turn to the left away from the light. To explain the behavior of an insect that approaches a light Loeb would say that the nerve pathways must be crossed so that the output of the eye receiving the stronger stimuli is transmitted to the legs on the opposite side of the body [*see illustration on page 76*]. Loeb contended that, regardless of whether the nerve pathways were crossed or uncrossed, they were separate and noninteracting. According to the tonus theory, locomotion in a straight line ultimately resulted when the pulling force exerted by opposite legs was in balance.

Loeb held to this theory with a tenacity that was surprising in view of the fact that contemporary physiologists could demonstrate extensive central-nervous-system interactions among nerve signals arising from various sense organs. The principal evidence for nerve integration was being provided by Sherrington, who was then at the University of Liverpool. Working with dogs and other mammals, Sherrington concentrated on the simplest forms of behavior: reflexes. In order to investigate reflexes without interference from the animal's higher brain centers Sherrington would transect the upper spinal cord (producing a "spinal animal") or the lower brain regions (producing a "decerebrate animal"). In this way he was able to study those reflexes of the trunk and legs whose controlling neurons (nerve cells) reside primarily in the relatively simple spinal cord.

Through his careful analysis of reflex behavior Sherrington discovered several ways in which nerve impulses, conducted along different sensory neurons, in-

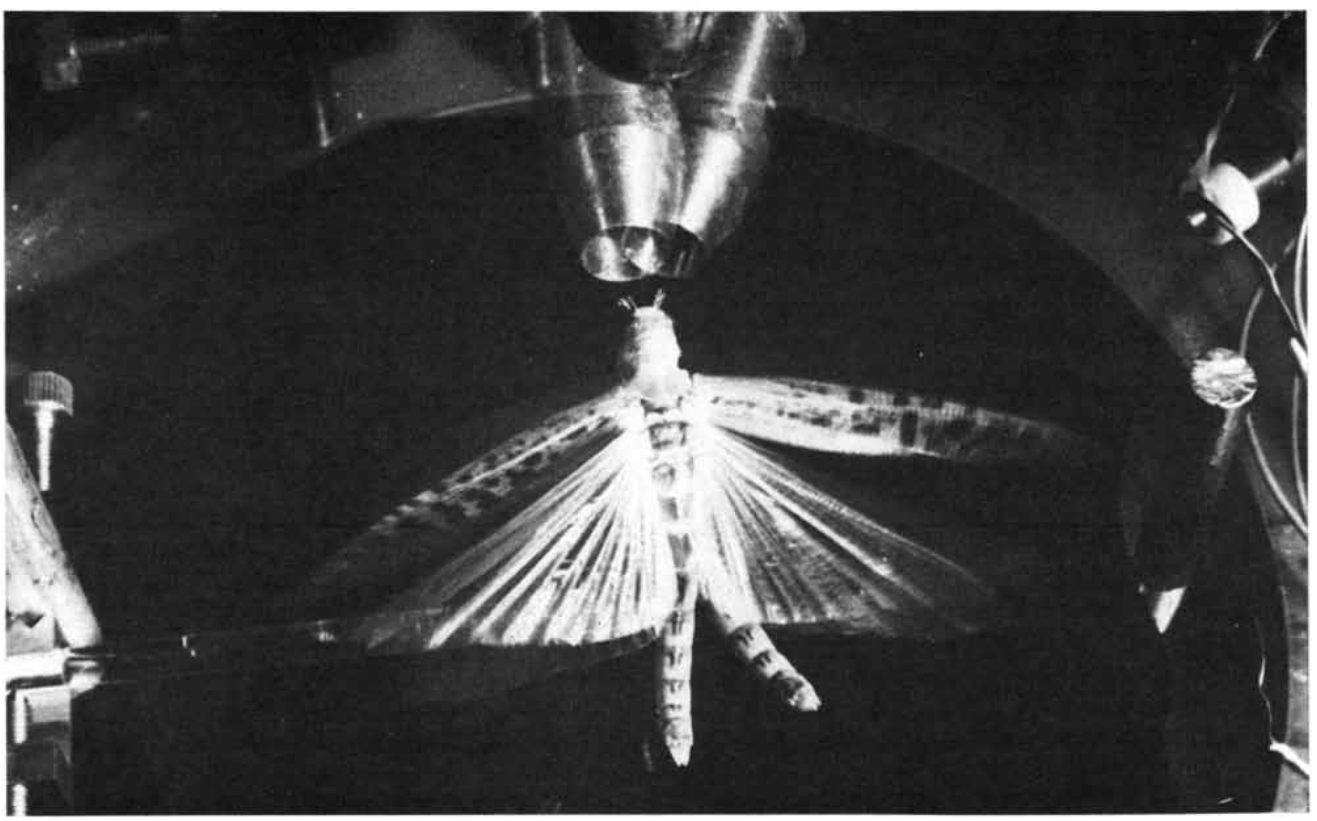
teract within the central nervous system. On indirect evidence he suggested that such impulses interact at synapses: the specialized loci of close apposition where impulses pass from one neuron to the next in a chain. (The existence of synapses was still in some doubt at the time, but there is now a wealth of anatomical and physiological evidence supporting Sherrington's view.)

One principle Sherrington discovered was "reciprocal innervation," an arrangement in which the excitation of the motor neurons innervating one group of muscles (resulting in their contraction) is accompanied by simultaneous inhibition of the motor neurons innervating an antagonistic group of muscles (those whose contraction would cause the opposite movement). The situation being reciprocal, the excitation of the second group of motor neurons is accompanied by the inhibition of the first. This arrangement, which eliminates the need for a muscle to physically overpower its antagonist, is one of the commonest neural patterns. Reciprocal interactions are also found in parts of the body that are some distance apart. Sherrington found, for example, that a spinal dog would withdraw a leg that received a sharp poke and would simultaneously brace the opposite leg to assume the weight removed from the withdrawn leg. What happened, purely by reflex, was that impulses from the spinal cord excited motor neurons in the adductor muscles of the poked leg and inhibited motor neurons in the abductor muscles; simultaneously impulses inhibited the adductor muscles of the opposite leg and excited the abductors [*see illustration on page 77*].

In one famous experiment on spinal dogs Sherrington found that stimulating any point within a broad area of the back by an "electric flea" (a point source of current) evoked rhythmic scratching movements of the hind leg on that side. Currents too weak to evoke immediate scratching could do so after prolonged stimulation, owing to "temporal summation" of stimuli at synapses within the central nervous system. Moreover, several stimuli individually too weak to evoke scratching could do so if they were applied simultaneously to separate regions on one side of the back, through the synaptic process of "spatial summation." Finally, if two stimuli, each by itself adequate to evoke scratching, were applied simultaneously to opposite sides of the back, only one side at a time would be scratched. The two legs often alternated in scratching sequences, suggesting that neurons controlling the scratching reflexes of opposite legs inhibit each other synaptically.

Ultimately Sherrington's experiments destroyed Loeb's hypothesis that sensory-motor pathways were completely separate by demonstrating in many ways that inputs from various receptors interact with one another centrally both to excite and to inhibit muscular responses. These studies suggested that the balancing of bilateral inputs is achieved by synaptic rather than mechanical events. The spinal cord was viewed as the seat of many of these synaptic interactions, with the brain exercising a number of complex controls over the basic spinal events. Although Sherrington was careful not to overgeneralize from his findings, less circumspect workers soon adopted the extreme position that most animal behavior consists of reflexes.

The current view leaves room for a wide range of reflex behavior but emphasizes that neuronal networks can



DESERT LOCUST (*Schistocerca gregaria*) is held by wax in front of a miniature wind tunnel in the author's laboratory at Cornell University. The wind tunnel is pivoted so that the airflow can be directed at the insect's head from various angles. In this double exposure the wind direction has been rotated from directly head on to a position some 25 degrees to the right, thus simulating what happens when the locust unintentionally yaws to the left. The locust's nervous system spontaneously initiates the corrections needed for reorientation. The corrections include twisting of the wings, rotation of the head, swinging of the legs and turning of the abdomen.

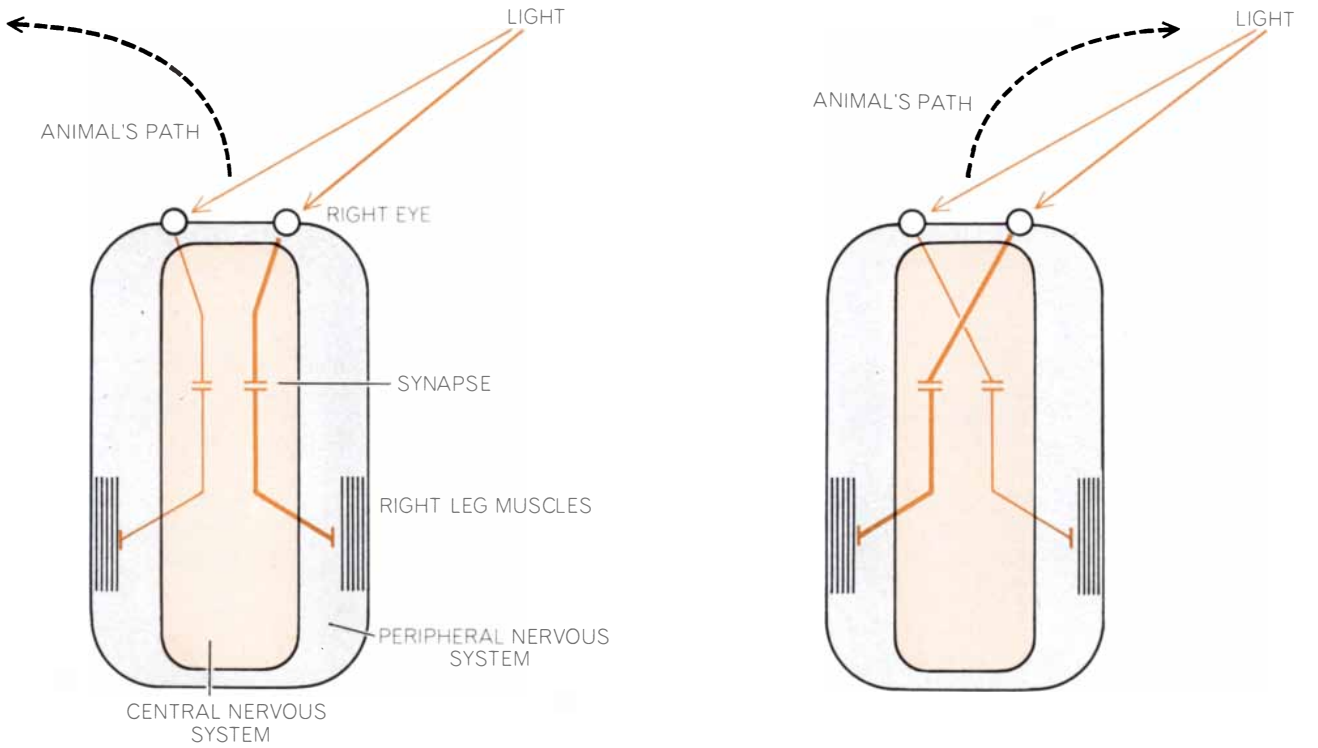
serve many functions independently of sensory inputs. Early evidence for this hypothesis was provided by the late Donald M. Wilson, who worked first at the University of California at Berkeley and then at Stanford University. Wilson demonstrated in locusts that the nerve impulses that rhythmically activate the wing muscles result from neural processing that takes place within thoracic ganglia: pools of neurons in the central nervous system between the neck and the abdomen. Without requiring sensory inputs, the thoracic ganglia provide the nerve impulses that make the elevator and depressor muscles for the locust's wings contract in rhythmic alternation; the impulses are also timed to make the back pair of wings beat slightly in advance of the front pair. Wilson termed the responsible network of neural interconnections the "central flight motor" [see "The Flight-Control System of the Locust," by Donald M. Wilson; SCIENTIFIC AMERICAN, May, 1968].

Only three naturally occurring sensory stimuli are specifically known to modify the motor's basic wingbeat pattern [*see illustration at top right on page 78*]. The flow of wind over receptor hairs on the locust's head can turn the motor on and maintain the flight rhythm. Contact stimulation of receptors on the locust's feet (tarsal receptors) can turn the motor off, ensuring that the wings stop beating when the insect alights. Lastly, the stretch imparted to one receptor cell on each wing during every upstroke acts as a stimulus to increase the frequency of the wingbeat.

It hardly needs saying that a locust deprived of a rich sensory inflow would have difficulty foraging, migrating and propagating. The central flight motor must be provided with sensory inputs to enable the locust to steer toward or away from objects and simply to maintain proper body orientation during flight. Several workers have shown that a variety of natural stimuli impose fine controls on the basic flight motor by modifying the number of impulses per wingbeat conducted to two or three tiny muscles in each forewing that control the twist of the wing. The angle of twist is one of the important controls for changing the direction of flight and for achieving stabilization.

My own work has dealt mainly with the way the locust achieves orientation with respect to yaw so that it does not turn crosswise to its line of flight. Some years ago Torkel Weis-Fogh, working in Copenhagen, had shown that a flying locust can monitor the angle of the wind passing over its body surface and that it can use this angle as a measure of its yaw orientation. As a locust flies it creates a wind over its surface (the "relative wind") just as we do when we run. If the locust makes a yawing error, its momentum will carry it for an instant along its original straight path, although it is now facing sideways. This causes a change in the angle of the relative wind that is equal and opposite to the yaw angle [*see bottom illustration on page 78*]. Weis-Fogh and I have both shown that the wind-receptor hairs on the locust's head sense the wind angle and evoke the necessary flight changes to restore proper orientation.

One thing I wished to determine was the precise body movements locusts employ in yaw reorientation. Using wax as an adhesive to hold the locust atop a fixed rod, I tethered my subjects at the mouth of a miniature wind tunnel that I could pivot rapidly in an arc in front of the insect's head. In this way I could expose the wind-receptor hairs to changes in wind angle mimicking the changes



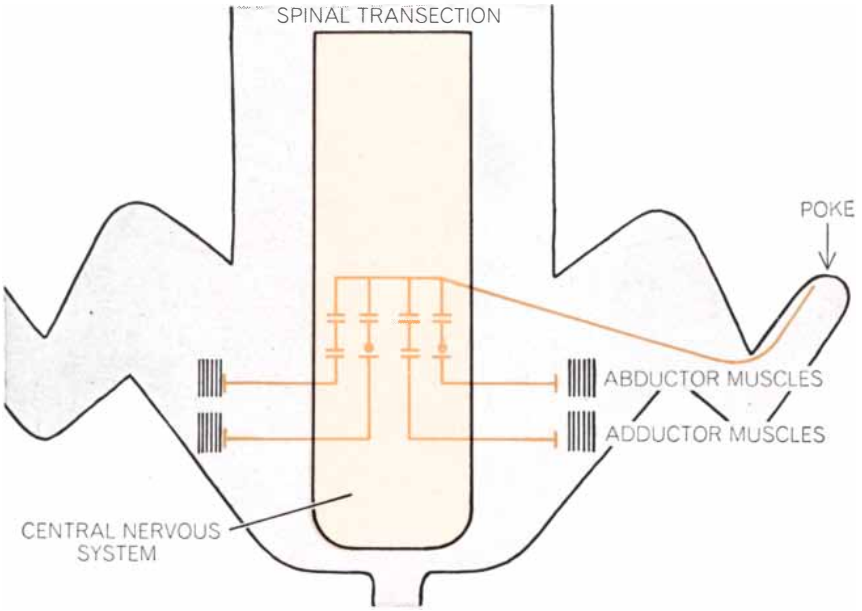
ONE EXPLANATION OF RESPONSE TO LIGHT in animals invoked the hypothesis, developed early in the 20th century by Jacques Loeb, that the nervous system is designed to maintain the bilateral symmetry of stimuli acting on the organism. On this view a "cockroach" (*left*) turns away from a light source because the eye nearer the light receives a stronger stimulus. This leads in turn to a greater flow of nerve impulses to leg muscles on that side, resulting in a lengthened stride that causes the animal to veer away from the light. To explain how a hypothetical light-loving cockroach (*right*) might be attracted to light one assumes, on the Loeb hypothesis, that the nerve pathways from the animal's eyes to its legs are crossed somewhere in the central nervous system. Now the greater flow of impulses to muscles in the legs opposite the light source lengthens the stride on that side and causes animal to turn toward light.

the locust would experience during a yaw in free flight. As soon as the airflow was turned on, the locust's wings would begin beating at their normal frequency and would usually continue as long as the insect's feet were not allowed to make contact with a surface. The experiments were conducted in total darkness to make sure that the response to artificial yaws was evoked by changes in the wind angle alone and not by the sight of the moving wind tunnel. With a stroboscopic flash camera I recorded the precise movements evoked when the airflow was quickly pivoted from head on to various angles [*see illustration on page 75*].

The pictures show that if the airflow mimics a yaw to the left, which requires the locust to initiate a correcting turn to the right, the following four responses all occur within about a third of a second. On each downstroke the right forewing twists down and/or the left forewing twists up. As viewed from the top, the insect's abdomen swings to the right as if it were a rudder. The second and third pairs of legs also swing to the right. The head rotates clockwise as if it were the nose of an airplane banking to the right. If the wind source is pivoted to mimic a yaw in the opposite direction, the corrective responses are reproduced in the opposite sense.

It is difficult to study the wing and leg movements quantitatively, but I have measured the simpler responses of the abdomen and the head. Both clearly increase in magnitude if the angle of the simulated yaw is increased or if one increases the speed with which the wind direction is turned through a given angle. If the wind direction is altered at a constant speed, the movements of the head and abdomen increase linearly with the simulated yaw angle over a range of from less than five degrees to more than 20 degrees in either direction [*see illustrations on page 79*].

Weis-Fogh had found that the wind-receptor hairs on the locust's head induce yaw reorientation, but he had not determined whether or not it is the hairs that induce the specific body movements that underlie the reorientation. It was not difficult to show that the receptors that give rise to the abdominal movements are located somewhere on the head. I carefully cut through the locust's neck cuticle, sparing the nerve pathways from the head to the flight motor, and rotated the insect's head 180 degrees. I then fixed the head firmly with wax to the thorax so that the head's left and right sides were functionally reversed. If the receptors that controlled the abdominal response were located on the head, one would now expect the response to be exactly reversed when the locust was held in front of the wind tunnel and subjected to a simulated yaw. This proved to be the case. I then covered selected regions of the locust's head with wax and verified that it was indeed the wind-receptor hairs and not some other structure that provided the stimuli for the movement of the abdomen.

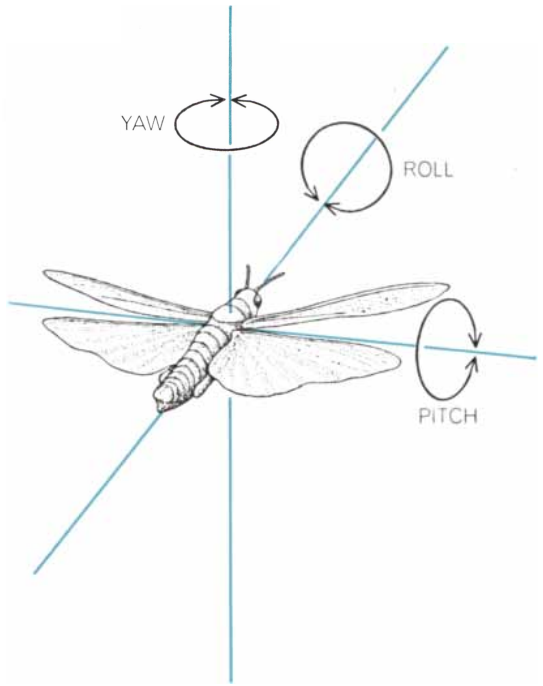


LEG-WITHDRAWAL REFLEX IN DOG was elucidated early in this century by Sir Charles Sherrington, who studied the behavior of animals whose spinal cord had been cut at the base of the brain. He proposed that impulses from peripheral sensory nerve cells interact within the central nervous system at synapses, or nerve junctions, where impulses are either transmitted from one neuron to the next or are prevented from passing. Synapses that block impulses, indicated here by a knob, are called inhibitory. Synapses in a dog's spinal cord are so arranged that if one of the dog's paws is sharply poked, the impulses reaching the dog's spinal cord excite motor neurons to the adductor muscles that cause the paw to be withdrawn; at the same time the abductors in the opposite leg are activated to carry the weight shifted to that leg. By "reciprocal innervation" the antagonistic muscles that would oppose the activated muscles are simultaneously inhibited from contracting.

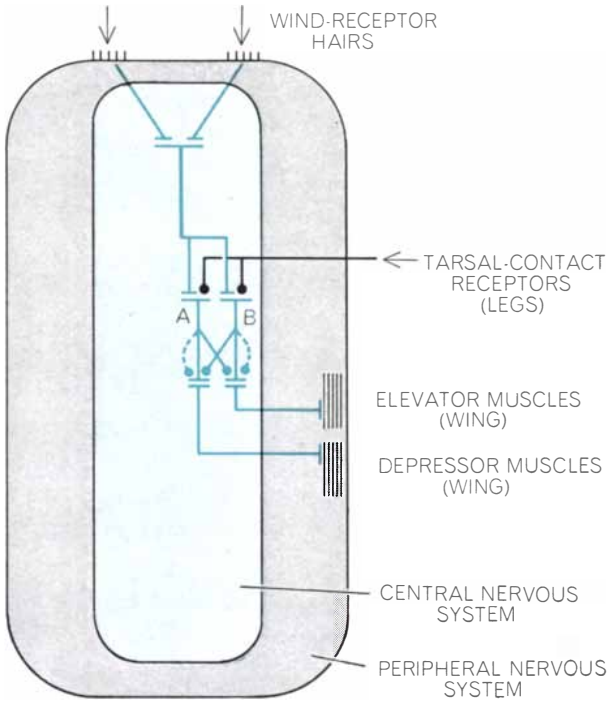
Although the movements of the legs, the wings and the abdomen that I observed in my experiments could be understood as dynamic responses that enable the locust to correct for yaw, it was not immediately clear why the locust should also rotate its head. Some earlier studies conducted by Peter T. Haskell and Leslie Goodman of the Anti-Locust Research Centre in London provided a clue. They had discovered a second group of hair receptors in the locust's cervical, or neck, region that respond not to wind but to the deflection of their shafts that occurs when the locust rotates its head around its body's long axis.

Goodman demonstrated that this sensory information is essential for roll stabilization. Since insects lack the special gravity receptors found in most other animals, they must employ other means to control roll and pitch, both of which involve orientations with respect to gravity. Goodman found that a locust responds to an incorrect roll attitude by rotating its head until its eyes are aligned parallel to the horizon or until the upper part of each eye is maximally stimulated by light from above. Either of these visual alignments will normally return the insect's head to the proper roll orientation. The cervical hair receptors then sense that the head has been rotated and evoke reflexive movements that will realign the thorax and abdomen with the head.

One of my first hypotheses was that the same body movements the locust uses to correct errors in roll might be used to correct errors in yaw. Conceivably in both cases the movements of the wings, the legs and the abdomen might be driven by head rotation acting through the cervical receptors. This turned out not to be the case, since the head rotation is actually the last (by a fraction of a second) of the four correcting movements to occur. I also found



THREE ORIENTATION AXES specify the rotations that must be controlled if any flying machine is to achieve stable flight. The locust is provided with sensors for detecting rotations around each of the three axes and the means for making corrections. The author has been concerned chiefly with the responses that preserve yaw orientation.

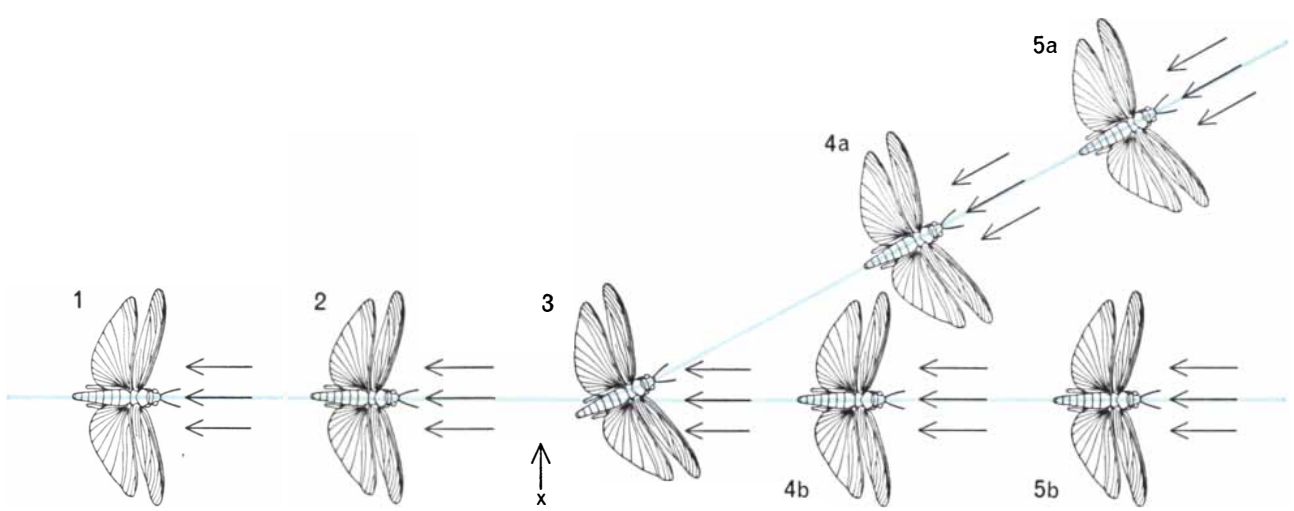


MODEL OF LOCUST "FLIGHT MOTOR" includes the nerve pathways and synapses depicted in color. The motor supplies the nerve impulses that rhythmically and alternately contract the elevator and depressor muscles of the locust's wings (only one of four wings is shown). Inputs from wind-receptor hairs on the locust's head can start the motor; inputs from contacts on the locust's feet can turn it off. Interneurons *A* and *B* are thought to have mutually inhibitory synapses to ensure that elevator and depressor muscles contract alternately. Each is also shown with a self-inhibiting process (*broken lines*) that provides a brief recovery period after each firing burst. Only a portion of the flight motor is depicted.

that the other responses take place even when the locust's head is fixed to the thorax with wax so that it cannot move.

My next idea was that perhaps the head rotation and consequent cervical receptor stimulus provide an alternative, or "fail safe," mechanism for correcting yaws, employing the same movements used in roll correction. These movements might be superimposed on primary yaw responses directly evoked by the head hairs, or they might occur only if the primary mechanism fails. The idea seemed attractive because my experiments had shown that yaw-correcting movements of the abdomen are not perfectly reliable.

To test this fail-safe notion I set the wind source so that it blew directly at the locust and then, employing a delicate torsion device, I rotated the insect's



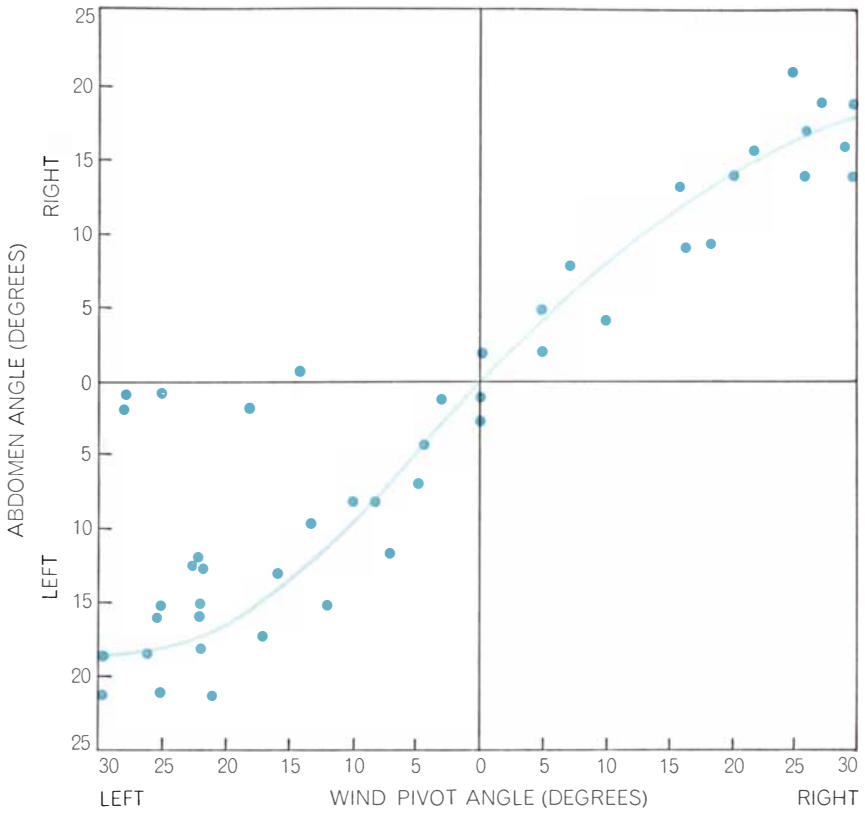
LOCUST FLIGHT PATH is represented with an inadvertent yaw at point *x*. The arrows indicate the direction of the relative wind as sensed by the hairlike wind receptors on the locust's head. At position No. 3, following the yaw, the relative wind strikes the receptors from the right side. If the yaw is uncorrected, the insect will follow the new flight path indicated by *4a* and *5a*. If the yaw is corrected by appropriate movements of the wings, head, legs and abdomen, the insect will regain its original flight path (*4b*, *5b*).

head around the stationary thorax, thereby mimicking the way the head normally moves when the wind is pivoted to one side. Almost every time I rotated the head counterclockwise (as viewed from behind) I observed a slight movement of the abdomen and legs to the left, the same movements normally evoked by pivoting the wind to the left. The movements had only about a third the magnitude of those normally evoked by a change in wind direction, and they took three times as long to complete.

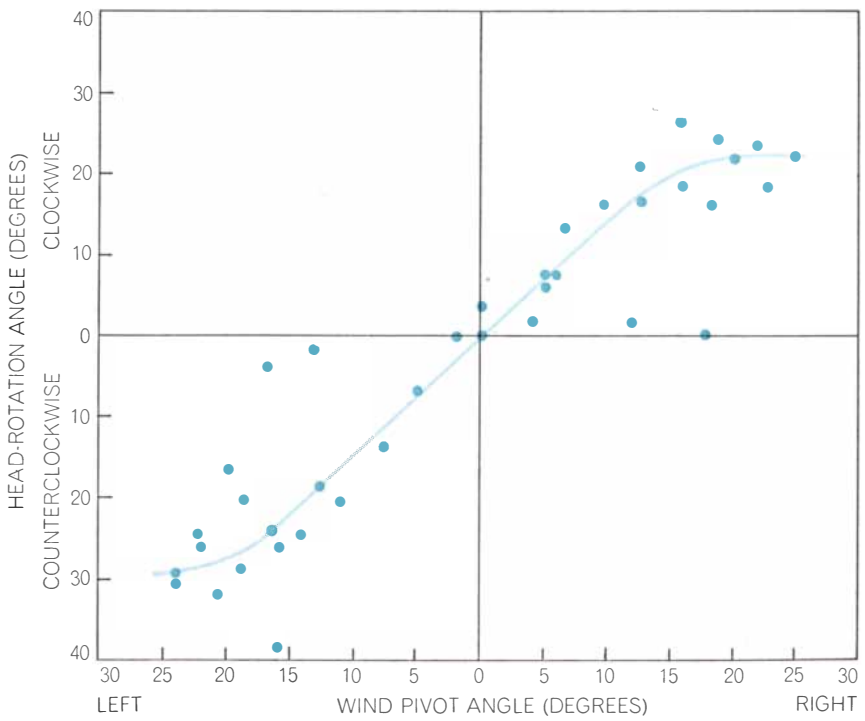
In order to ascertain whether these movements resulted from stimulation of the cervical hairs or from other receptors still unknown I removed the cervical receptors from several locusts and tested the behavior of the insects over a period of several days. The locusts whose receptors had been excised no longer moved their abdomen or legs when I forcibly rotated their head, although other aspects of their flight were normal, suggesting that the cervical receptors are indeed responsible for the movement. Even without the cervical receptors, however, the locusts still responded normally to changes in wind direction, owing to the stimulation of the head hairs.

Thus it appears that locusts have available two different yaw-correcting strategies: (1) a rapid change in wing twist, abdomen position and leg position controlled by wind-sensitive hairs on the head and (2) a slower, subtler movement of the same general character evoked by cervical receptors. One can speculate that the slower set of responses arose in the course of evolution expressly to correct roll and that at some point a neural pathway was introduced that tied the roll-correction movements into the faster set of responses that had evolved independently to correct yaw.

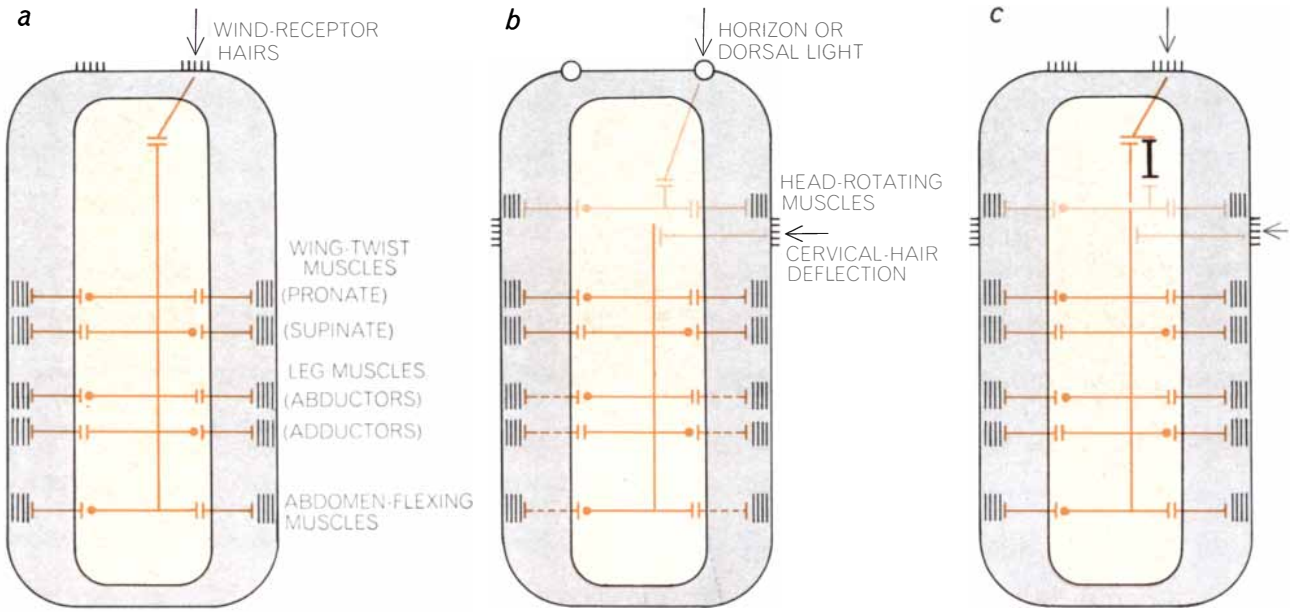
Another question that interested me is how the locust coordinates the movements of eight different body parts (the head, the abdomen, the one set of wings and four of the six legs) into a single integrated response. The answer has proved elusive. Careful comparison of head movements and abdomen movements shows that there is not much proportionality between them. If the head fails to respond to a particular change in wind direction, the abdomen may show a normal response, and vice versa. In fact, one can remove the locust's entire abdomen, together with its legs and wings, and find that the head still responds to changes in wind angle. It therefore seems that the change in wind angle, indicating a yaw, is integrated



MOVEMENTS OF LOCUST'S ABDOMEN induced by simulated yaws are plotted for various changes in wind direction. As part of the yaw-correcting response the abdomen swings in the same direction as the wind shift. Changes in wind angle up to about 20 degrees in either direction evoke a nearly linear change in the angle between abdomen and thorax.



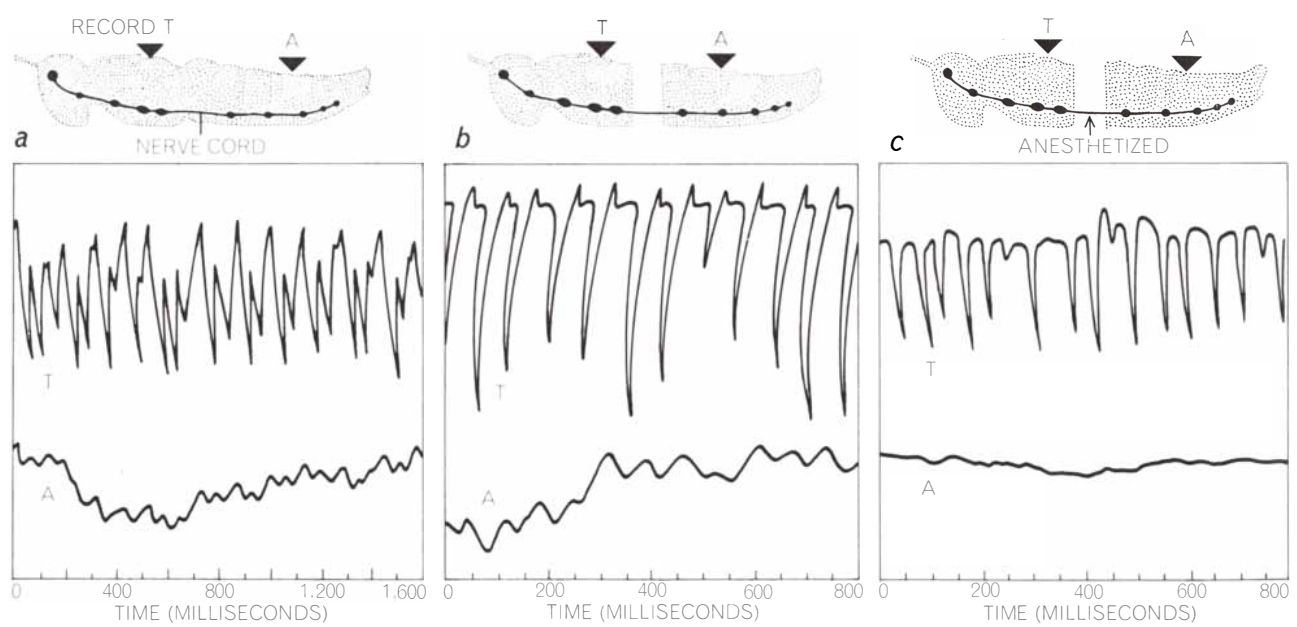
ROTATION OF LOCUST'S HEAD induced by yaws either to left or right also shows a nearly linear response for wind shifts up to about 20 degrees. For example, in response to a yaw to the left (wind pivot to the right) the head rotates clockwise as seen from the rear.



HYPOTHETICAL EVOLUTION of the locust's complete yaw-correcting system from a fast-yaw system and an independent-roll system is depicted in three stages. The fast-yaw-correcting system (*a*) may have consisted of a single sensory pathway from wind-receptor hairs to a central control mechanism that evoked rapid changes in the twist of the forewings together with changes in the position of the legs and abdomen. The roll-stabilizing system (*b*) may have consisted of additional pathways (*lines in light color*) that provided an input evoked by the deflection of hairs on the neck and an output to muscles that rotate the head. (Broken lines in *b* indicate pathways not directly demonstrated.) The complete yaw-correcting system (*c*) may have evolved by the addition of a single neural connection (*heavy black line*) that enables nerve impulses from wind-receptor hairs to rotate the head in response to a yaw.

somewhere in the locust's central nervous system, and is followed by independent motor commands to the wings, legs, abdomen and head.

Whatever the integrative mechanism is, it is clear that the movements associated with yaw correction occur only if the locust is flying. When, as often happens, a locust tethered in the miniature windstream spontaneously stops flying, it no longer responds to changes in wind direction. The moment flight spontaneously resumes, however, all the yaw responses can be evoked in full measure. Something about flying itself makes the locust's central nervous system "attentive" to the wind-angle information supplied by the head hairs. Locusts not in flight "ignore" the same information. Maija Hinkle and I have tried to find out what determines whether or not the lo-



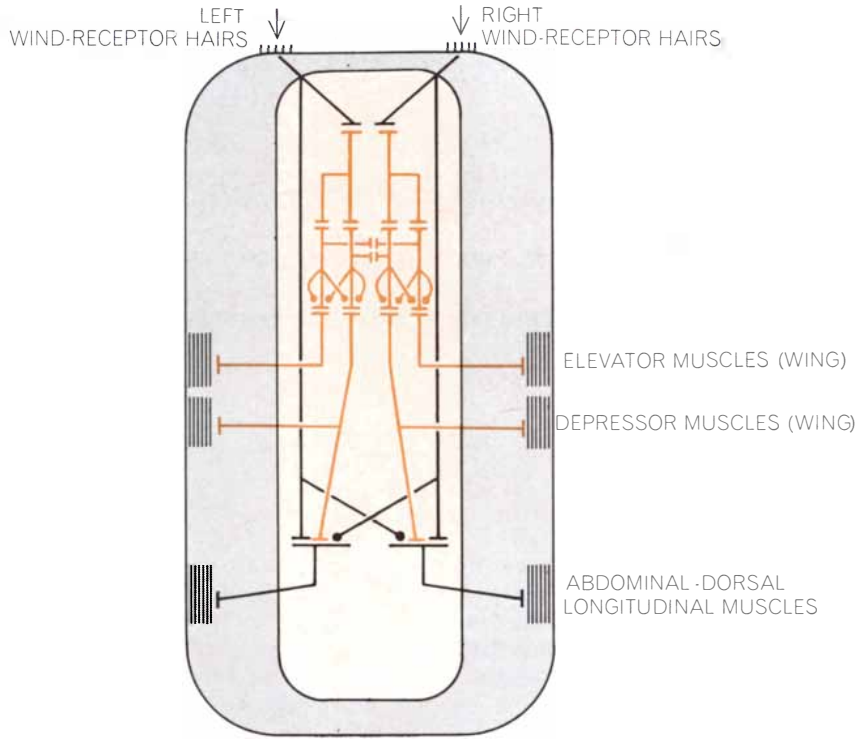
MOVEMENT OF LOCUST'S ABDOMEN AND THORAX during flight was recorded in the author's laboratory. During normal flight (*a*) rapid up-and-down vibrations of the abdomen (*record A*) are locked in phase with vibrations of the thorax (*record T*), which are largely responsible for wing movements. If the abdomen and the thorax are separated so that only the soft pliable nerve cord connects the two, both rhythms persist and remain in phase (*b*). This suggests that abdominal vibrations are independently generated by abdominal muscles and are not due simply to a mechanical linkage between the abdomen and the thorax. If the nerve-cord bridge between the thorax and the abdomen is anesthetized (*c*), the abdominal rhythm is eliminated but not the thoracic rhythm.

cust's central nervous system responds to wind-angle messages. To simplify the problem we focused our attention on the simple, rudder-like response of the abdomen to changes in wind angle.

One of our first observations was that during flight the locust's abdomen is constantly active, making slight but rapid vibrations in the vertical plane. Suspecting that these vibrations might somehow be related to the yaw-correcting movements of the abdomen, we wished to see whether the vibrations resulted from contractions of the abdominal muscles or merely reflected the abdomen's mechanical linkage to the thorax, whose vibrations are largely responsible for the wing movements. Mechanical and electrical recordings made simultaneously from the abdomen and the thorax showed that the two vibrations have the same frequency and are locked in phase [*see bottom illustration on opposite page*]. When we separated the two body sections, leaving only the pliable nerve cord as a bridge between the two, both rhythms remained; moreover, their frequencies were still equal and in phase, suggesting that the abdominal muscles play a part in generating the abdominal vibration. When we anesthetized the nerve-cord bridge, we eliminated the abdominal rhythm but not the thoracic one, indicating that neurons somewhere in the thorax or the head conduct the rhythmic impulses to the abdominal muscles.

Electrical recordings of the entire motor nerve to the main lifting muscle of a single abdominal segment (the dorsal longitudinal muscle, located in bilateral pairs in each segment) reveal that during flight the motor neurons conduct impulses to these muscles in bursts that are synchronous with similar bursts in the motor nerves leading to the forewings. Contractions of the dorsal longitudinal muscles on both sides of each abdominal segment produce the abdomen's vibrations during flight.

The way the muscles are attached to the abdomen suggested to us that the normal up-and-down vibration of the abdomen during flight is caused by bilaterally equal contractions of the muscles and that a rudder-like movement would result if the contractions were stronger on one side than on the other. This hypothesis was borne out by our experiments. A head-on wind evoked approximately equal bursts in the left and right nerves, whereas when the wind was shifted to either side of the locust's head, there was an increase in the number of motor-neuron impulses per wingbeat on that side accompanied by a decrease in the strength of the bursts on the other side. Pivoting the wind as much as 20 degrees would often completely suppress the bursts in the opposite nerve. (The burst frequency and its phase relation to the wingbeat were unaffected by wind angle.) It appeared, then, that wind-angle messages operated during flight by superimposing a fine control on the ongoing impulse bursts conducted to the abdominal muscles throughout flight.



LOCUST'S ABDOMINAL VIBRATIONS are traced to the locust's central flight motor (*color*) and its interactions with sensory inputs from the wind-angle receptors. The abdomen is raised and lowered during flight by dorsal longitudinal muscles whose inputs are arranged as shown. Normally the muscles on both sides are equally stimulated. If, however, the wind receptors should detect a yaw to the left, the additional impulses from the right wind-receptor hairs would tend to inhibit the dorsal longitudinal contractions on the insect's left side and increase those on the right, thus deflecting the abdomen to the right.

Where do these ongoing rhythmic bursts in the abdominal nerves originate? Are they commands emanating directly from the central flight motor, or are they the result of feedback from some sense receptors stimulated rhythmically by the wingbeat? (Such rhythmic sensory stimulation had been shown to occur in at least one stretch receptor on each wing.) In order to make the distinction between central and sensory origins, we isolated almost the entire central nervous system of the locust by cutting all peripheral nerves of the abdomen and thorax, thereby interrupting all possible sources of rhythmic sensory feedback. When we applied wind to the locust's head to excite the central flight motor, recordings made from the cut central stumps of both an abdominal and a thoracic motor nerve showed synchronous, fully normal rhythms in both. Pivoting the wind produced the same effects on abdominal bursts as when all the peripheral connections of the abdomen and the thorax were intact.

Thus the rhythmic abdominal bursts appear to originate in the central flight motor. We can now see that this central nervous network performs an even wider function than Wilson had originally determined. The rhythmic central flight motor not only patterns the wingbeat but also raises the abdomen into flight posture through rapid vertical vibrations, thereby priming the abdominal neurons to respond to fine controls from wind-receptor inputs [*see illustration above*]. The priming function of the flight motor seems to explain why the locust is "attentive" to yaw-angle information only during flight. The same priming function probably applies to the motor elements that control the twist of the forewings and perhaps also to the elements that control the leg and head movements.

The Control of Short-Term Memory

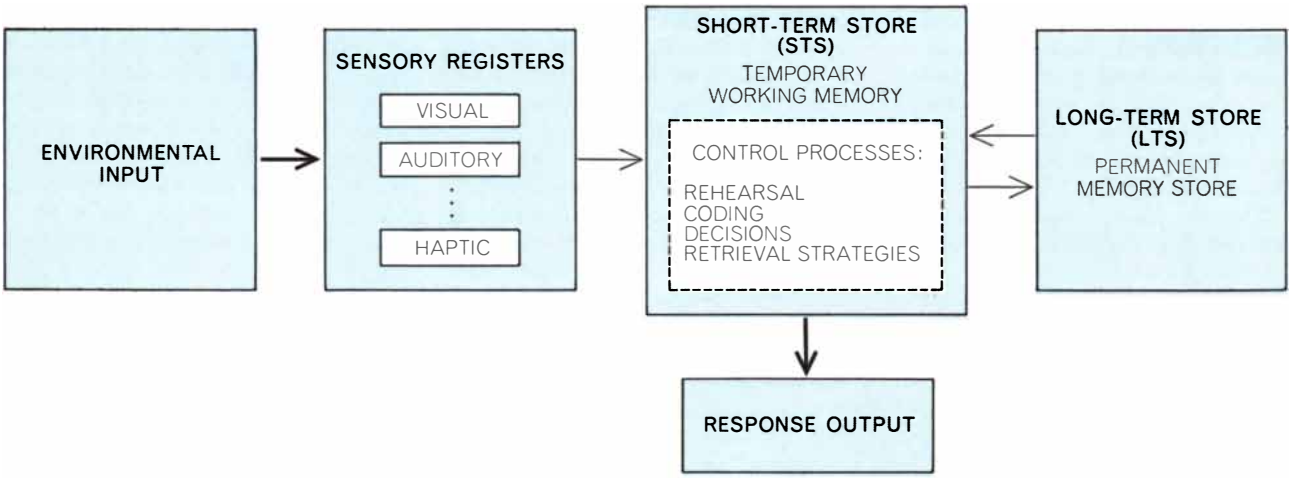
Memory has two components: short-term and long-term. Control processes such as "rehearsal" are essential to the transfer of information from the short-term store to the long-term one

by Richard C. Atkinson and Richard M. Shiffrin

The notion that the system by which information is stored in memory and retrieved from it can be divided into two components dates back to the 19th century. Theories distinguishing between two different kinds of memory were proposed by the English associationists James Mill and John Stuart Mill and by such early experimental psychologists as Wilhelm Wundt and Ernst Meumann in Germany and William James in the U.S. Reflecting on their own mental processes, they discerned a clear difference between thoughts currently in consciousness and thoughts that could be brought to consciousness only after a search of memory that was often laborious. (For example, the sentence you are reading is in your current awareness; the name of the baseball team that won the 1968 World Series may be in your memory, but to retrieve it takes some effort, and you may not be able to retrieve it at all.)

The two-component concept of memory was intuitively attractive, and yet it was largely discarded when psychology turned to behaviorism, which emphasized research on animals rather than humans. The distinction between short-term memory and long-term memory received little further consideration until the 1950's, when such psychologists as Donald E. Broadbent in England, D. O. Hebb in Canada and George A. Miller in the U.S. reintroduced it [see "Information and Memory," by George A. Miller; SCIENTIFIC AMERICAN, August, 1956]. The concurrent development of computer models of behavior and of mathematical psychology accelerated the growth of interest in the two-process viewpoint, which is now undergoing considerable theoretical development and is the subject of a large research effort. In particular, the short-term memory system, or short-term store (STS), has been given a position of pivotal importance. That is because the processes carried out in the short-term store are under the immediate control of the subject and govern the flow of information in the memory system; they can be called into play at the subject's discretion, with enormous consequences for performance.

Some control processes are used in many situations by everyone and others are used only in special circumstances. "Rehearsal" is an overt or covert repetition of information—as in remembering a telephone number until it can be writ-



INFORMATION FLOW through the memory system is conceived of as beginning with the processing of environmental inputs in sensory registers (receptors plus internal elements) and entry into the short-term store (STS). While it remains there the information may be copied into the long-term store (LTS), and associated information that is in the long-term store may be activated and entered into the short-term store. If a triangle is seen, for example, the name "triangle" may be called up. Control processes in the short-term store affect these transfers into and out of the long-term store and govern learning, retrieval of information and forgetting.

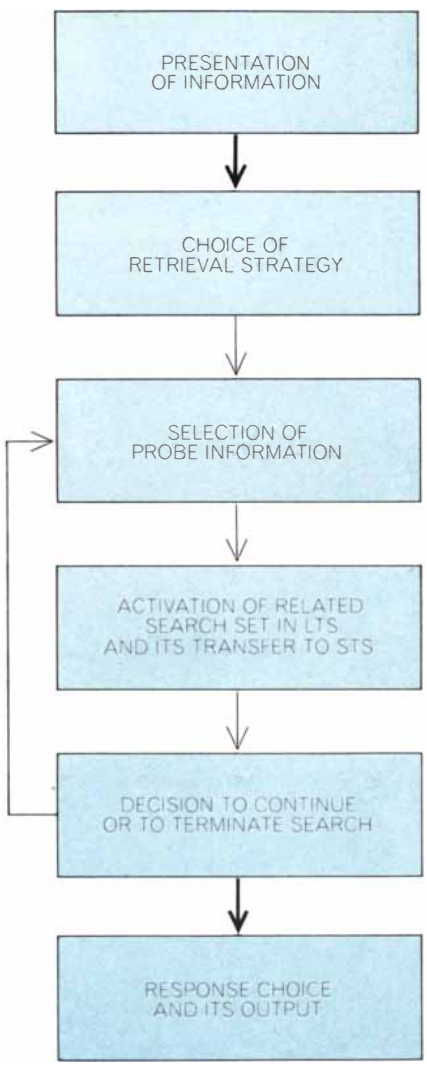
ten down, remembering the names of a group of people to whom one has just been introduced or copying a passage from a book. "Coding" refers to a class of control processes in which the information to be remembered is put in a context of additional, easily retrievable information, such as a mnemonic phrase or sentence. "Imaging" is a control process in which verbal information is remembered through visual images; for example, Cicero suggested learning long lists (or speeches) by placing each member of the list in a visual representation of successive rooms of a well-known building. There are other control processes, including decision rules, organizational schemes, retrieval strategies and problem-solving techniques; some of them will be encountered in this article. The point to keep in mind is the optional nature of control processes. In contrast to permanent structural components of the memory system, the control processes are selected at the subject's discretion; they may vary not only with different tasks but also from one encounter with the same task to the next.

We believe that the overall memory system is best described in terms of the flow of information into and out of short-term storage and the subject's control of that flow, and this conception has been central to our experimental and theoretical investigation of memory. All phases of memory are assumed to consist of small units of information that are associatively related. A set of closely interrelated information units is termed an image or a trace. Note that "image" does not necessarily imply a visual representation; if the letter-number pair *TKM–4* is presented for memory, the image that is stored might include the size of the card on which the pair is printed, the type of print, the sound of the various symbols, the semantic codes and numerous other units of information.

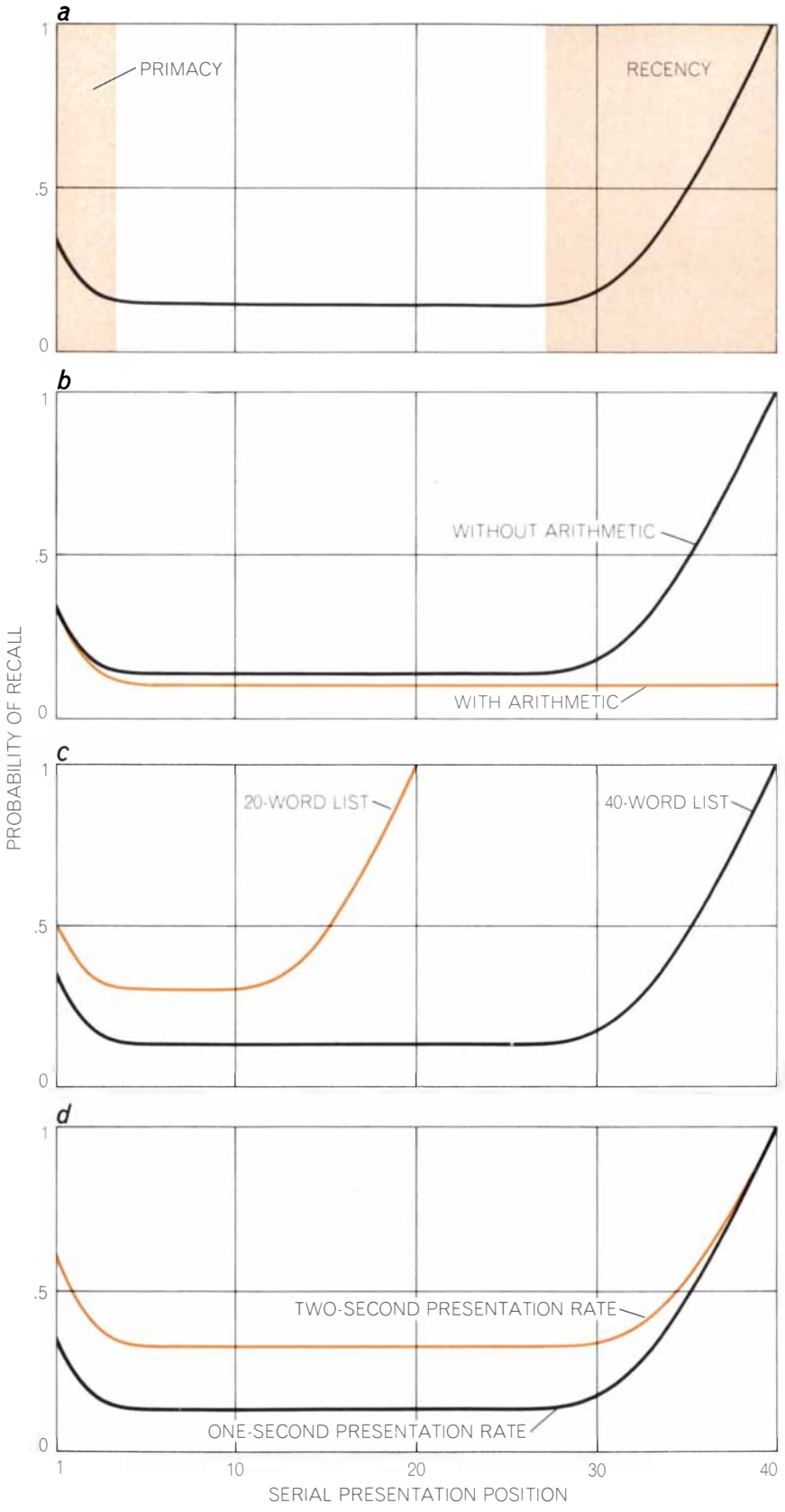
Information from the environment is accepted and processed by the various sensory systems and is entered into the short-term store, where it remains for a period of time that is usually under the control of the subject. By rehearsing one or more items the subject can keep them in the short-term store, but the number that can be maintained in this way is strictly limited; most people can maintain seven to nine digits, for example. Once an image is lost from the short-term store it cannot thereafter be recovered from it. While information resides in short-term storage it may be copied into the long-term store (LTS), which is assumed to be a relatively permanent memory from which information is not lost. While an image is in short-term storage, closely related information in the long-term store is activated and entered in the short-term store too. Information entering the short-term store from the sensory systems comes from a specific modality—visual, auditory or whatever—but associations from the long-term store in all modalities are activated to join it. For instance, an item may be presented visually, but immediately after input its verbal "name" and associated meanings will be activated from the long-term store and placed in the short-term one [*see illustration on opposite page*].

Our account of short-term and long-term storage does not require that the two stores necessarily be in different parts of the brain or involve different physiological structures. One might consider the short-term store simply as being a temporary activation of some portion of the long-term store. In our thinking we tend to equate the short-term store with "consciousness," that is, the thoughts and information of which we are currently aware can be considered part of the contents of the short-term store. (Such a statement lies in the realm of phenomenology and cannot be verified scientifically, but thinking of the short-term store in this way may help the reader to conceptualize the system.) Because consciousness is equated with the short-term store and because control processes are centered in and act through it, the short-term store is considered a working memory: a system in which decisions are made, problems are solved and information flow is directed. Retrieval of information from short-term storage is quite fast and accurate. Experiments by Saul Sternberg of the Bell Telephone Laboratories and by others have shown that the retrieval time for information in short-term storage such as letters and numbers ranges from 10 to 30 milliseconds per character.

The retrieval of information from long-term storage is considerably more complicated. So much information is contained in the long-term store that the major problem is finding access to some small subset of the information that contains the desired image, just as one must find a particular book in a library before it can be scanned for the desired information. We propose that the subject activates a likely subset of information, places it in the short-term store and then scans that store for the desired image. The image may not be present in the current subset, and so the retrieval process becomes a search in which various subsets are successively activated and scanned [*see illustration below*]. On the basis of the information presented to him the subject selects the appropriate "probe information" and places it in the short-term store. A "search set," or subset of information in the long-term store closely associated with the probe, is then activated and put in the short-term store. The subject selects from the search set some image, which is then examined. The information extracted from the selected image is utilized for a decision: has the desired information



RETRIEVAL from the long-term store requires a choice of strategy and selection of certain information as a "probe" that is placed in the short-term store. The probe activates a "search set" of information in the long-term store. The search set is placed in the short-term store and is examined for the desired information. If it is not found, search is halted or recycled with new probe.



PROBABILITY OF RECALL in free-recall experiments varies in a characteristic way with an item's serial position in a list: a "primacy effect" and a "recency effect" are apparent (*a*). If an arithmetic task is interpolated between presentation and recall, the recency effect disappears (*b*). Words in long lists are recalled less well than words in short lists (*c*). Slower presentation also results in better recall (*d*). The curves are idealized ones based on experiments by James W. Dees, Bennet Murdock, Leo Postman and Murray Glanzer.

been found? If so, the search is terminated.

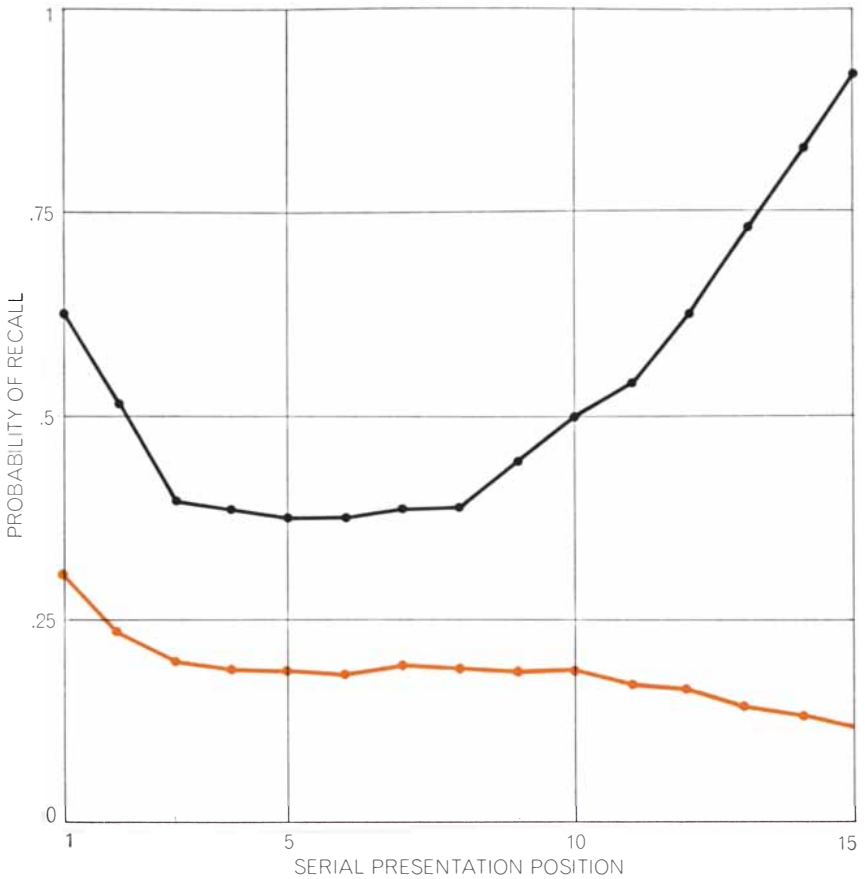
If the information has not been found, the subject may decide that continuation is unlikely to be productive or he may decide to continue. If he does, he begins the next cycle of the search by again selecting a probe, which may or may not be the same probe used in the preceding cycle depending on the subject's strategy. For example, a subject asked to search for states of the U.S. starting with the letter *M* may do so by generating states at random and checking their first letter (in which case the same probe information can be used in each search cycle), or he may generate successive states in a regular geographic order (in which case the probe information is systematically changed from one cycle to the next). It can be shown that strategies in which the probe information is systematically changed will result more often in successful retrieval but will take longer than alternative "random" strategies. (Note that the Freudian concept of repressed memories can be considered as being an inability of the subject to generate an appropriate probe.)

This portrayal of the memory system almost entirely in terms of the operations of the short-term store is quite intentional. In our view information storage and retrieval are best described in terms of the flow of information through the short-term store and in terms of the subject's control of the flow. One of the most important of these control processes is rehearsal. Through overt or covert repetition of information, rehearsal either increases the momentary strength of information in the short-term store or otherwise delays its loss. Rehearsal can be shown not only to maintain information in short-term storage but also to control transfer from the short-term store to the long-term one. We shall present several experiments concerned with an analysis of the rehearsal process.

The research in question involves a memory paradigm known as "free recall," which is similar to the task you face when you are asked to name the people present at the last large party you went to. In the typical experimental procedure a list of random items, usually common English words, is presented to the subject one at a time. Later the subject attempts to recall as many words as possible in any order. Many psychologists have worked on free recall, with major research efforts carried out by

Bennet Murdock of the University of Toronto, Endel Tulving of Yale University and Murray Glanzer of New York University. The result of principal interest is the probability of recalling each item in a list as a function of its place in the list, or "serial-presentation position." Plotting this function yields a *U*-shaped curve [*see "a" in illustration on opposite page*]. The increased probability of recall for the first few words in the list is called the primacy effect; the large increase for the last eight to 12 words is called the recency effect. There is considerable evidence that the recency effect is due to retrieval from short-term storage and that the earlier portions of the serial-position curve reflect retrieval from long-term storage only. In one experimental procedure the subject is required to carry out a difficult arithmetic task for 30 seconds immediately following presentation of the list and then is asked to recall. One can assume that the arithmetic task causes the loss of all the words in short-term storage, so that recall reflects retrieval from long-term storage only. The recency effect is eliminated when this experiment is performed; the earlier portions of the serial-position curve are unaffected [*b*]. If variables that influence the long-term store but not the short-term one are manipulated, the recency portion of the serial-position curve should be relatively unaffected, whereas the earlier portions of the curve should show changes. One such variable is the number of words in the presented list. A word in a longer list is less likely to be recalled, but the recency effect is quite unaffected by list length [*c*]. Similarly, increases in the rate of presentation decrease the likelihood of recalling words preceding the recency region but leave the recency effect largely unchanged [*d*].

In free recall experiments many lists are usually presented in a session. If the subject is asked at the end of the session to recall all the words presented during the session, we would expect his recall to reflect retrieval from long-term storage only. The probability of recalling words as a function of their serial position within each list can be plotted for end-of-session recall and compared with the serial-position curve for recall immediately following presentation [*see illustration on this page*]. For the delayed-recall curve the primacy effect remains, but the recency effect is eliminated, as predicted. In summary, the recency region appears to reflect retrieval from both short-term and long-term storage whereas the serial-position curve preceding the recency region reflects retrieval from long-term storage only.



EFFECT OF DELAY is tested by asking subjects to recall at the end of a session all words from the entire session, and then plotting probability of recall against serial position within each list. An experiment by Fergus Craik compares immediate recall (*black*) with delayed recall (*color*). The delayed-recall curve emphasizes transitory nature of recency effect.

In 1965, at a conference sponsored by the New York Academy of Sciences, we put forward a mathematical model explaining these and other effects in terms of a rehearsal process. The model assumed that in a free-recall task the subject sets up a rehearsal buffer in the short-term store that can hold only a fixed number of items. At the start of the presentation of a list the buffer is empty; successive items are entered until the buffer is filled. Thereafter, as each new item enters the rehearsal buffer it replaces one of the items already there. (Which item is replaced depends on a number of psychological factors, but in the model the decision is approximated by a random process.) The items that are still being rehearsed in the short-term store when the last item is presented are the ones that are immediately recalled by the subject, giving rise to the recency effect. The transfer of information from the short-term to the long-term store is postulated to be a function of the length of time an item resides in the rehearsal buffer; the longer the time period, the more rehearsal the item receives and therefore the greater the transfer of information to long-term storage. Since items presented first in a list enter an empty or partly empty rehearsal buffer, they remain longer than later items and consequently receive additional rehearsal. This extra rehearsal causes more transfer of information to long-term storage for the first items, giving rise to the primacy effect.

This rehearsal model was given a formal mathematical statement and was fitted to a wide array of experiments, and it provided an excellent quantitative account of a great many results in free recall, including those discussed in this article. A more direct confirmation of the model has recently been provided by Dewey Rundus of Stanford University. He carried out free-recall experiments in which subjects rehearsed aloud during list presentation. This overt rehearsal was tape-recorded and was com-

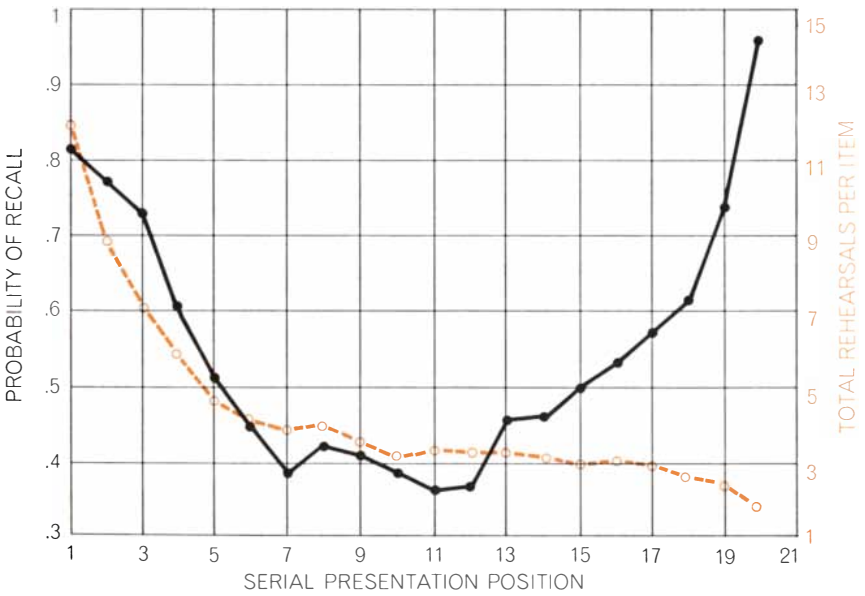
pared with the recall results. The number of different words contained in the "rehearsal set" (the items overtly rehearsed between successive presentations) was one after the first word was presented and then rose until the fourth word; from the fourth word on the number of different words in the rehearsal set remained fairly constant (averaging about 3.3) until the end of the list. The subjects almost always reported the members of the most recent rehearsal set when the list ended and recall began. A close correspondence is evident between the number of rehearsals and the recall probability for words preceding the recency effect; in the recency region, however, a sharp disparity occurs [*see illustrations below*]. The hypothesis that long-term storage is a function of the number of rehearsals can be checked in other ways. The recall probability for a word preceding the recency region was plotted as a function of the number of rehearsals received by that word; the result was an almost linear, sharply increasing function. And words presented in the middle of the list given the same number of rehearsals as the first item presented had the same recall probability as that first item.

With efficacy of rehearsal established both for storing information in the long-term store and for maintaining information in the short-term store, we did an experiment in which the subjects' rehearsal was manipulated directly. Our subjects were trained to engage in one of two types of rehearsal. In the first (a one-item rehearsal set) the most recently presented item was rehearsed exactly three times before presentation of the next item; no other items were rehearsed. In the second (a three-item rehearsal set) the subject rehearsed the three most recently presented items once each before presentation of the next item, so that the first rehearsal set contained three rehearsals of the first word, the second rehearsal set contained two rehearsals of the second word and one rehearsal of the first word, and all subsequent sets contained one rehearsal of each of the three most recent items [*see illustrations on opposite page*].

When only one item is rehearsed at a time, each item receives an identical number of rehearsals and the primacy effect disappears, as predicted. Note that the recency effect appears for items preceding the last item even though the last item is the only one in the last rehearsal set. This indicates that even when items are dropped from rehearsal, it takes an additional period of time for them to be completely lost from short-term storage. The curve for the three-item rehearsal condition shows the effect also. The last rehearsal set contains the last three items presented and these are recalled perfectly, but a recency effect is still seen for items preceding these three. It should also be noted that a primacy effect occurs in the three-rehearsal condition. This was predicted because the first item received a total of five rehearsals rather than three. A delayed-recall test for all words was given at the end of the experimental session. The data confirmed that long-term-store retrieval closely parallels the number of rehearsals given an item during presentation, for both rehearsal schemes.

ITEM PRESENTED	ITEMS REHEARSED (REHEARSAL SET)
1 REACTION	REACTION, REACTION, REACTION, REACTION
2 HOOF	HOOF, REACTION, HOOF, REACTION
3 BLESSING	BLESSING, HOOF, REACTION
4 RESEARCH	RESEARCH, REACTION, HOOF, RESEARCH
5 CANDY	CANDY, HOOF, RESEARCH, REACTION
6 HARDSHIP	HARDSHIP, HOOF, HARDSHIP, HOOF
7 KINDNESS	KINDNESS, CANDY, HARDSHIP, HOOF
8 NONSENSE	NONSENSE, KINDNESS, CANDY, HARDSHIP
⋮	⋮
20 CELLAR	CELLAR, ALCOHOL, MISERY, CELLAR

OVERT-REHEARSAL experiment by Dewey Rundus shows the effect of rehearsal on transfer into long-term storage. The subject rehearses aloud. A partial listing of items rehearsed in one instance shows typical result: early items receive more rehearsals than later items.



EFFECT OF REHEARSAL is demonstrated by comparison of an item's probability of recall (*black*) with the total number of rehearsals item receives (*color*). The two are related in regions reflecting retrieval from long-term storage (preceding recency region). That is, long-term storage efficacy depends on number of rehearsals and is reflected in retrieval.

These results strongly implicate rehearsal in the maintenance of information in the short-term store and the transfer of that information to the long-term system. The question then arises: What are the forgetting and transfer characteristics of the short-term store in the absence of rehearsal? One can control rehearsal experimentally by blocking it with a difficult verbal task such as arithmetic. For example, Lloyd R. Peterson and Margaret Peterson of Indiana University [see "Short-Term Memory," by Lloyd R. Peterson; SCIENTIFIC AMERICAN, July, 1966] presented a set of three letters (a trigram) to be remembered; the subject next engaged in a period of arithmetic and then was asked to recall as many letters of the trigram

as possible. When the probability of recall is plotted as a function of the duration of the arithmetic task, the loss observed over time is similar to that of the recency effect in free recall [*see top illustration on next page*]. Short-term-store loss caused by an arithmetic task, then, is similar to loss from short-term storage caused by a series of intervening words to be remembered. The flat portion of the curve reflects the retrieval of the trigram from long-term storage alone and the earlier portions of the curve represent retrieval from both short-term and long-term storage; the loss of the trigram from short-term storage is represented by the decreasing probability of recall prior to the asymptote.

Does the forgetting observed during arithmetic reflect an automatic decay of short-term storage that occurs inevitably in the absence of rehearsal or is the intervening activity the cause of the loss? There is evidence that the amount of new material introduced between presentation and test is a much more important determinant of loss from short-term storage than simply the elapsed time between presentation and test. This finding is subject to at least two explanations. The first holds that the activity intervening between presentation and test is the *direct* cause of an item's loss from short-term storage. The second explanation proposes that the rate of intervening activity merely affects the number of rehearsals that can be given the item to be remembered and thus *indirectly* determines the rate of loss.

It has recently become possible to choose between these two explanations of loss from the short-term store. Judith Reitman of the University of Michigan substituted a signal-detection task for the arithmetic task in the Petersons' procedure. The task consisted in responding whenever a weak tone was heard against a continuous background of "white" noise. Surprisingly, no loss from short-term storage was observed after 15 seconds of the task, even though subjects reported no rehearsal during the signal detection. This suggests that loss from the short-term store is due to the type of interference during the intervening interval: signal detection does not cause loss but verbal arithmetic does. Another important issue that could potentially be resolved with the Reitman procedure concerns the transfer of information from the short-term to the long-term store: Does transfer occur only at initial presentation and at subsequent rehearsals, or does it occur throughout the pe-

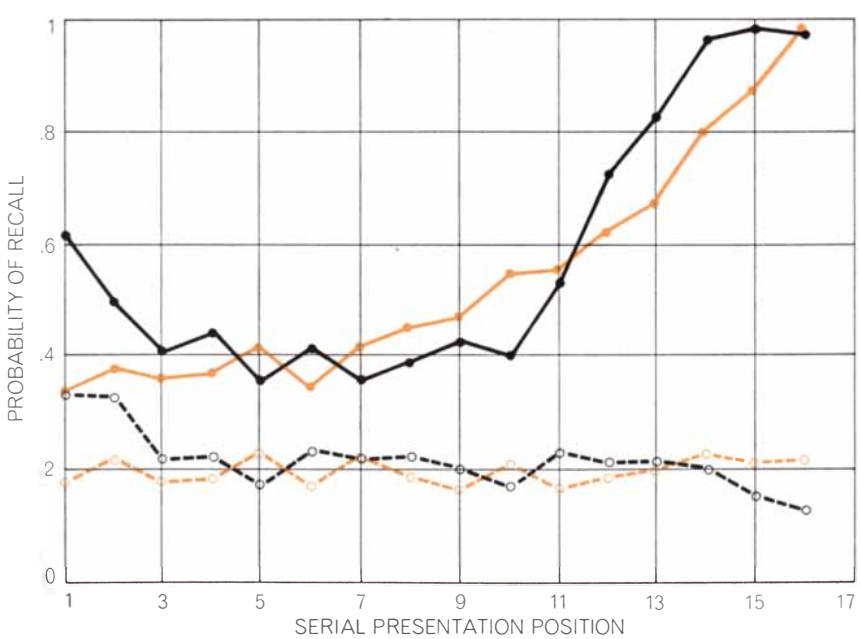
ONE-ITEM REHEARSAL SCHEME

SERIAL POSITION	ITEM PRESENTED	ITEMS REHEARSED	TOTAL REHEARSALS PER ITEM
1	A	AAA	3
2	B	BBB	3
3	C	CCC	3
4	D	DDD	3
5	E	EEE	3
6	F	FFF	3
.	.	.	.
.	.	.	.
.	.	.	.
14	N	NNN	3
15	O	OOO	3
16	P	PPP	3

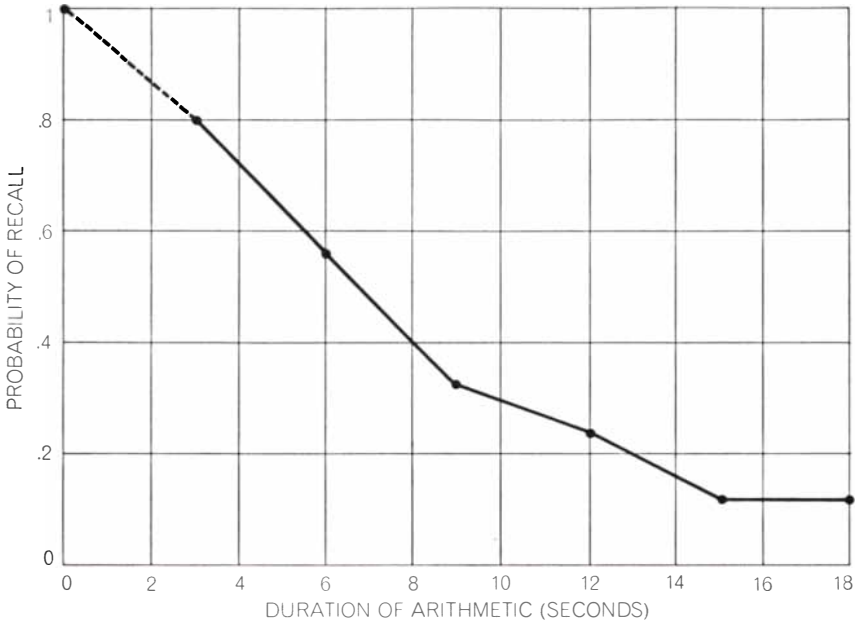
THREE-ITEM REHEARSAL SCHEME

SERIAL POSITION	ITEM PRESENTED	ITEMS REHEARSED	TOTAL REHEARSALS PER ITEM
1	A	AAA	5
2	B	BBA	4
3	C	CBA	3
4	D	DCB	3
5	E	EDC	3
6	F	FED	3
.	.	.	.
.	.	.	.
.	.	.	.
14	N	NML	3
15	O	ONM	2
16	P	PON	1

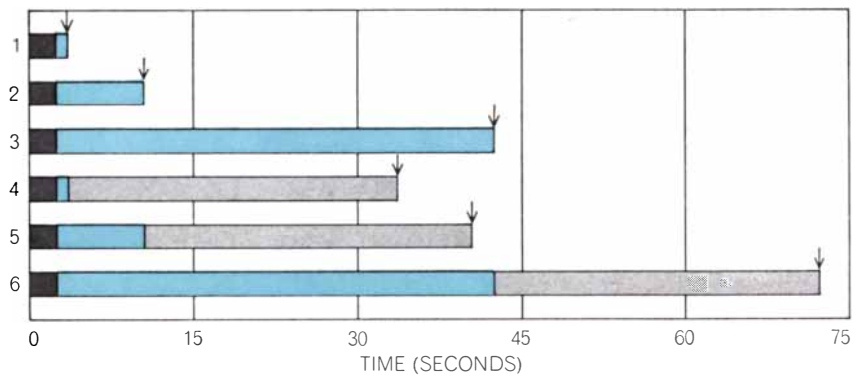
NUMBER OF REHEARSALS is controlled with two schemes. In one (*top*) only the current item is rehearsed and all items have three rehearsals. In the other (*bottom*) the latest three items are rehearsed; early ones have extra rehearsals. (Letters represent words.)



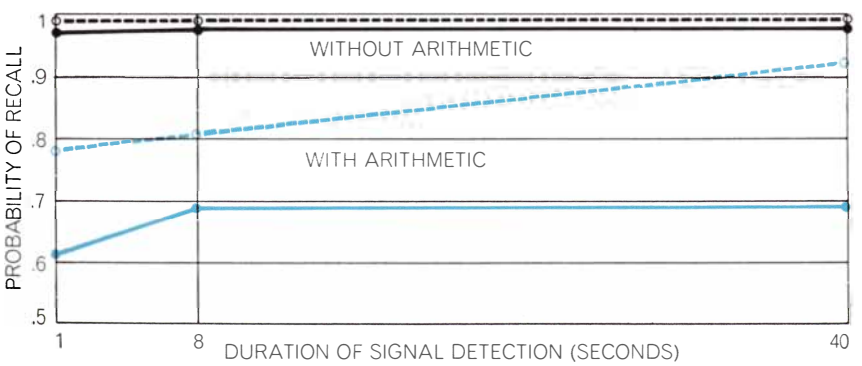
PRIMACY EFFECT disappears with one-item rehearsal (*color*), in which all items have equal rehearsal, but remains with three-item rehearsal (*black*). Recency effect is pronounced for both schemes in immediate recall (*solid lines*). Curves for delayed recall (*broken lines*), which reflect only retrieval from long-term storage, parallel the number of rehearsals.



ARITHMETIC TASK before recall reduces the probability of recall. Lloyd R. Peterson and Margaret Peterson charted recall probability against duration of arithmetic. The probability falls off with duration until it levels off when recall reflects retrieval from long-term storage alone. Does curve reflect only lack of rehearsal or also nature of intervening task?



TWO TASKS were combined in an experiment with these six conditions. Five consonants were presented for 2.5 seconds (*dark gray*), followed by a signal-detection task for one second, eight seconds or 40 seconds (*color*), followed in three cases by arithmetic (*light gray*). Then came the test (*arrows*). Rehearsal during detection was included in a control version.



NATURE OF TASKS is seen to have an effect. In the absence of arithmetic, signal detection leaves the short-term store virtually unaffected, with rehearsal (*broken black curve*) or without (*solid black*). Arithmetic, however, causes loss from the short-term store (*color*); decreased recall shown reflects retrieval from long-term store only. Retrieval improves with duration of signal detection if there is rehearsal, which increases transfer to the long-term store (*broken colored curve*) but not in the absence of rehearsal (*solid color*).

riod during which the information resides in the short-term store, regardless of rehearsals?

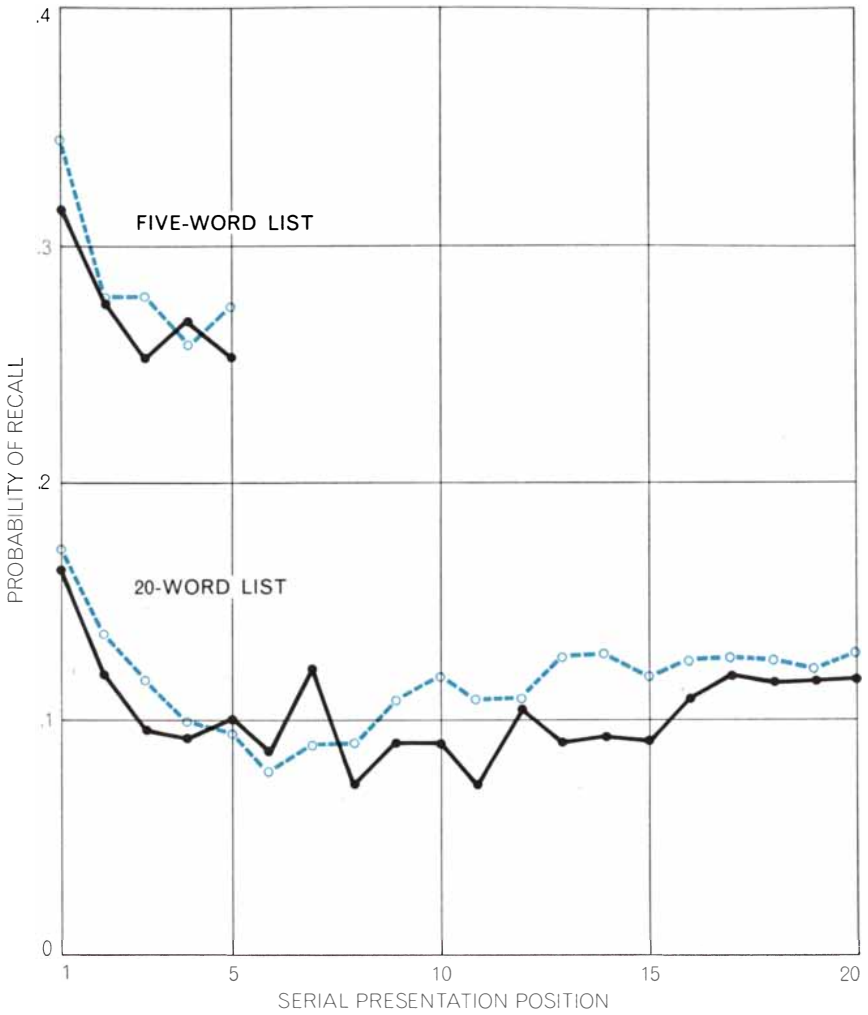
To answer these questions, the following experiment was carried out. A consonant pentagram (a set of five consonants, such as *QJXFK*) was presented for 2.5 seconds for the subject to memorize. This was followed by a signal-detection task in which pure tones were presented at random intervals against a continuous background of white noise. The subjects pressed a key whenever they thought they detected a tone. (The task proved to be difficult; only about three-fourths of the tones presented were correctly detected.) The signal-detection period lasted for either one second, eight seconds or 40 seconds, with tones sounded on the average every 2.5 seconds. In conditions 1, 2 and 3 the subjects were tested on the consonant pentagram immediately after the signal detection; in conditions 4, 5 and 6, however, they were required to carry out 30 seconds of difficult arithmetic following the signal detection before being tested [*see middle illustration at left*]. In order to increase the likelihood that rehearsal would not occur, we paid the subjects for performing well on signal detection and for doing their arithmetic accurately but not for their success in remembering letters. In addition they were instructed not to rehearse letters during signal detection or arithmetic. They reported afterward that they were not consciously aware of rehearsing. Because the question of rehearsal is quite important, we nevertheless went on to do an additional control experiment in which all the same conditions applied but the subjects were told to rehearse the pentagram aloud following each detection of a tone.

The results indicate that arithmetic causes the pentagram information to be lost from the short-term store but that in the absence of the arithmetic the signal-detection task alone causes no loss [*see bottom illustration at left*]. What then does produce forgetting from the short-term store? It is not just the analysis of any information input, since signal detection is a difficult information-processing task but causes no forgetting. And time alone causes no noticeable forgetting. Yet verbal information (arithmetic) does cause a large loss. Mrs. Reitman's conclusion appears to be correct: forgetting is caused by the entry into the short-term store of other, similar information.

What about the effect of rehearsal? In the arithmetic situation performance improves if subjects rehearse overtly

during the signal-detection period. Presumably the rehearsal transfers information about the pentagram to the long-term store; the additional transfer during the long signal-detection period is reflected in the retrieval scores, and the rehearsal curve rises. The no-rehearsal curve is horizontal over the last 32 seconds of signal detection, however, confirming that no rehearsal was occurring during that period. The fact that the lowest curve is flat over the last 32 seconds has important implications for transfer from the short-term store to the long-term. It indicates that essentially no transfer occurred during this period even though, as the results in the absence of arithmetic show, the trace remained in the short-term store. Hence the presence of a trace in the short-term store is alone not enough to result in transfer to the long-term store. Apparently transfer to the long-term system occurs primarily during or shortly after rehearsals. (The rise in the lowest curve over the first eight seconds may indicate that the transfer effects of a presentation or rehearsal take at least a few seconds to reach completion.)

The emphasis we have given to rote rehearsal should not imply that other control processes are of lesser importance. Although much evidence indicates that transfer from short-term storage to long-term is strongly dependent on rehearsals, effective later retrieval from long-term storage can be shown to be highly dependent on the type of information rehearsed. Coding is really the choosing of particular information to be rehearsed in the short-term store. In general, coding strategies consist in adding appropriately chosen information from long-term storage to a trace to be remembered and then rehearsing the entire complex in the short-term store. Suppose you are given (as is typical in memory experiments) the stimulus-response pair *HRM–4*; later *HRM* will be presented alone and you will be expected to respond "4." If you simply rehearse *HRM–4* several times, your ability to respond correctly later will probably not be high. Suppose, however, *HRM* reminds you of "homeroom" and you think of various aspects of your fourth-grade classroom. Your retrieval performance will be greatly enhanced. Why? First of all, the amount and range of information stored appears to be greater with coding than with rote rehearsal. Moreover, the coding operation provides a straightforward means by which you can gain access to an appropriate and small region of memory



LENGTH OF LIST rather than amount of "interference" governs recall probability. Subjects were asked to recall the list before the one just studied. Five-word lists (*top*) were recalled better than 20-word lists (*bottom*) whether they were followed by intervening lists of five words (*black*) or of 20 words (*color*). The data are averages from three experiments.

during retrieval. In the above example, when *HRM* is presented at the moment of test, you are likely to notice, just as during the initial presentation, that *HRM* is similar to "homeroom." You can then use "homeroom" (and the current temporal context) as a further probe and would almost certainly access "fourth grade" and so generate the correct response.

As the discussion of coding suggests, the key to retrieval is the selection of probe information that will activate an appropriate search set from the long-term store. Since in our view the long-term store is a relatively permanent repository, forgetting is assumed to result from an inadequate selection of probe information and a consequent failure of the retrieval process. There are two basic ways in which the probe selection may prove inadequate. First, the wrong probe may be selected. For instance, you might be asked to name the star of a particular motion picture. The name actually begins with *T* but you decide that it begins with *A* and include *A* in the probe information used to access the long-term store. As a result the correct name may not be included in the search set that is drawn into the short-term store and retrieval will not succeed.

Second, if the probe is such that an extremely large region of memory is accessed, then retrieval may fail even though the desired trace is included in the search set. For example, if you are asked to name a fruit that sounds like a word meaning "to look at," you might say "pear." If you are asked to name a living thing that sounds like a word meaning "to look at," the probability of your coming up with "pear" will be

greatly reduced. Again, you are more likely to remember a "John Smith" if you met him at a party with five other people than if there had been 20 people at the party. This effect can be explained on grounds other than a failure of memory search, however. It could be argued that more attention was given to "John Smith" at the smaller party. Or if the permanence of long-term storage is not accepted, it could be argued that the names of the many other people met at the larger party erode or destroy the memory trace for "John Smith." Are these objections reasonable? The John Smith example is analogous to the situation in free recall where words in long lists are less well recalled from long-term storage than words in short lists.

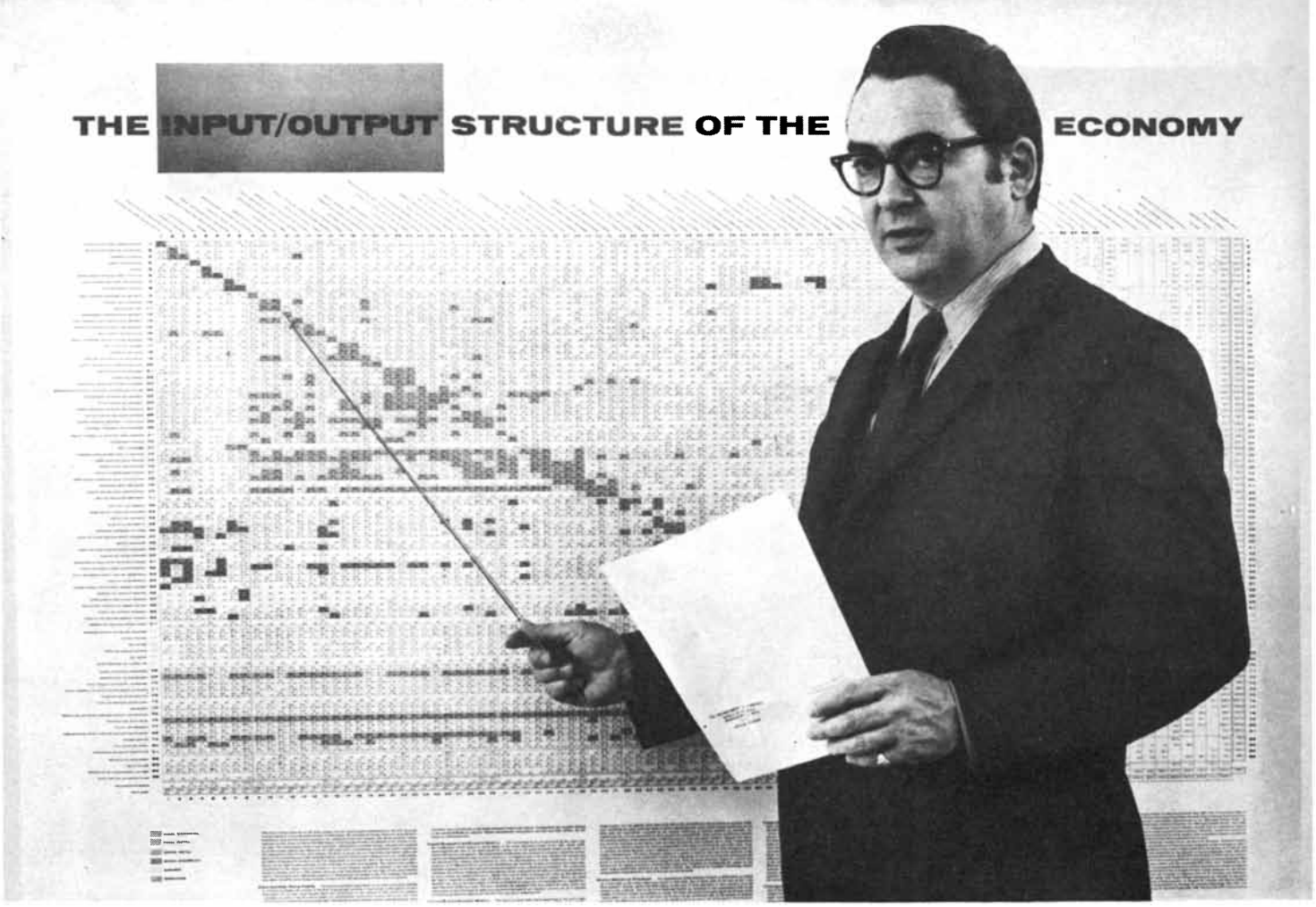
The problem, then, is to show that the list-length effect in free recall is dependent on the choice of probe information rather than on either the number of words intervening between presentation and recall or the differential storage given words in lists of different size. The second issue is disposed of rather easily: in many free-recall experiments that vary list length, the subjects do not know at the beginning of the list what the length of the list will be. It is therefore unlikely that they store different amounts of information for the first several words in lists of differing length. Nevertheless, as we pointed out, the first several words are recalled at different levels.

To dispose of the "interference" explanation, which implicates the number of words between presentation and recall, is more difficult. Until fairly recently, as a matter of fact, interference theories of forgetting have been predominant [see "Forgetting," by Benton J. Underwood, SCIENTIFIC AMERICAN, March, 1964, and "The Interference Theory of Forgetting," by John Ceraso, October, 1967]. In these theories forgetting has often been seen as a matter of erosion of the memory trace, usually by items presented following the item to be remembered but also by items preceding the item to be remembered. (The list-length effect might be explained in these terms, since the average item in a long list is preceded and followed by more items than the average item in a short list.) On the other hand, the retrieval model presented in this article assumes long-term storage to be permanent; it maintains that the strength of long-term traces is independent of list length and that forgetting results from the fact that the temporal-contextual probe cues used to access any given list tend to elicit a larger search set for longer lists, thereby producing less efficient retrieval.

In order to distinguish between the retrieval and the interference explanations, we presented lists of varying lengths and had the subject attempt to recall not the list just studied (as in the typical free-recall procedure) but the list before the last. This procedure makes it possible to separate the effect of the size of the list being recalled from the effect of the number of words intervening between presentation and recall. A large or a small list to be recalled can be followed by either a large or a small intervening list. The retrieval model predicts that recall probability will be dependent on the size of the list being recalled. The interference model predicts that performance will be largely determined by the number of words in the intervening list.

We used lists of five and of 20 words and presented them in four combinations: 5–5, 5–20, 20–5, 20–20; the first number gives the size of the list being recalled and the second number the size of the intervening list. One result is that there is no recency effect [*see illustration on preceding page*]. This would be expected since there is another list and another recall intervening between presentation and recall; the intervening activity causes the words in the tested list to be lost from short-term storage and so the curves represent retrieval from long-term storage only. The significant finding is that words in lists five words long are recalled much better than words in lists 20 words long, and the length of the intervening list has little, if any, effect. The retrieval model can predict these results only if a probe is available to access the requested list. It seems likely in this experiment that the subject has available at test appropriate cues (probably temporal in nature) to enable him to select probe information pertaining to the desired list. If the experimental procedure were changed so that the subject was asked to recall the 10th preceding list, then selection of an adequate probe would no longer be possible. The results demonstrate the importance of probe selection, a control process of the short-term store.

The model of memory we have described, which integrates the system around the operations of the short-term store, is not in any sense a final theory. As experimental techniques and mathematical models have become increasingly sophisticated, memory theory has undergone progressive changes, and there is no doubt that this trend will continue. We nevertheless think it is likely that the short-term store and its control processes will be found to be central.



WHAT DOES IT TAKE TO PRODUCE A $1000-BILLION GNP?

The Editors of SCIENTIFIC AMERICAN have prepared a wall chart, based upon the latest Federal input/output table, displaying the interindustry flows of raw materials, intermediate products and business services required to carry the U.S. economy to the benchmark Gross National Product of $1000 billion.

Input/output tables provide management, government administrators, economists and market analysts with a powerful new tool for forecasting and measuring the indirect as well as the direct interindustry relationships that structure our industrial economy.

This handsome and informative wall chart (70" x 46", in eight colors) offers a unique entry into the rapidly developing discipline of interindustry (or input/output) analysis. Based upon input/output tables issued by the Office of Business Economics of the U.S. Department of Commerce, the chart can be used as a teaching tool and for study of practical and theoretical questions about the U.S. economy.

The chart presents an interindustry matrix of 99 rows and 99 columns; each of the nearly 10,000 cells in the matrix shows (1) the direct input/output coefficient, (2) the "inverse" coefficient and (3) the interindustry dollar flow for a $1000-billion Gross National Product. The input/output coefficients as published by OBE have been recomputed by the Harvard Economic Research Project to reflect gross domestic output. The 370 sectors of the detailed tabulations have been selectively aggregated to 99 sectors to provide maximum feasible detail for the wall chart. Where the ratio of input to output exceeds 1/100, the cell is tinted in the color-code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

Offprints of five SCIENTIFIC AMERICAN articles on the technique of input/output analysis accompany the chart. The articles are:

Input/Output Economics
by Wassily W. Leontief

The Economic Effects of Disarmament
by Wassily W. Leontief and Marvin Hoffenberg

The Structure of Development
by Wassily W. Leontief

The Structure of the U.S. Economy
by Wassily W. Leontief

The Economics of Technological Change
by Anne P. Carter

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

I enclose $_____. (California residents please add sales tax.)
Send me, postpaid, _____ input/output wall charts at $10 each, plus the offprints listed.

Name_____

Company_____

Address_____

_____

New Models of the Real-Number Line

Recent developments in mathematical logic reveal that there are a number of alternative ways of defining the continuum, or connected number system, to include all the real numbers

by Lynn Arthur Steen

Virtually all of mathematics and much of science is based on the abstract concept of the real-number line: the continuum, or connected number system, that includes all the real numbers—zero, positive and negative integers, rational numbers (fractions) and irrational numbers (such as π)—but that excludes the "unreal," or complex, numbers—expressions containing the imaginary number $\sqrt{-1}$. The real-number continuum not only provides the natural setting for all the operations of arithmetic and calculus but also serves as our only intellectual model of time and (one-dimensional) space. The properties of the continuum were organized into a coherent axiomatic framework during the 19th century and have been accepted and promulgated with great conviction by most contemporary mathematicians.

Yet in spite of the present unanimity of opinion concerning the exact structure of the continuum several significant alternative systems have been discovered during the past 10 years. Although none of these new models reflects any logical flaw in the 19th-century theory, their very existence shows quite clearly that the epistemological foundation of mathematical analysis is far from settled.

Although mathematics is sometimes called a science, it is usually distinguished from science by its relative independence from empirical considerations. The intellectual models of science are judged by their ability to explain the observed properties of the universe, whereas those of mathematics are judged (by mathematicians) according to their consistency and beauty and (by scientists) according to their utility. Plato's description of mathematics as the discovery of the properties of objects in an ideal universe—the universe of Platonic ideals—has been the most enduring and popular philosophy of mathematics, since it provides mathematics with the (scientific) discipline of conforming to some kind of perceived reality together with the freedom to escape the bonds of empiricism.

Most working mathematicians, at least those not directly engaged in problems of mathematical logic, tend to be Platonists in the sense that they talk and act as if the abstract objects they study have some kind of enduring ideal existence. Standing in extreme opposition to the Platonists are the Formalists, who maintain that the essence of mathematics is not in its meaning but in its form.

The problem of the nature of the real-number line is viewed quite differently by adherents to these two schools of thought. A Formalist is likely to say that the real-number line is whatever we define it to be; if we have several competing definitions, then we will have several different real-number lines and mathematics will be enriched by their presence. A Platonist, on the other hand, would be inclined to wonder which of the different models represented the "real" real-number abstraction of the space and time continuum.

Like beauty, mathematics exists in the eye of the beholder. Unlike beauty, however, mathematics enjoys an unparalleled worldwide reputation for objectivity. If you labor to discover the properties of the real-number continuum and if I do likewise, we shall reach the same conclusions. Science has a similar objectivity. But whereas the objectivity of science is a rather plausible consequence of its conformity with reality, the objectivity of mathematics is harder to explain. Why should the ideal mathematical universe in your mind be the same as the one in my mind?

Two centuries ago Immanuel Kant attempted to answer this question by postulating the a priori existence in the human mind of a kernel of intuitive mathematical and geometrical truth. The considerable influence of Kant's philosophy reinforced the widely held view that the axioms of mathematics should be self-evident truths. Euclidean geometry became the archetype of mathematics, since it was a beautiful, useful theory created by logical deduction from certain (nearly) self-evident axioms. Indeed, Kant himself had asserted that the geometrical intuition of Euclidean space and of the time continuum was one of the a priori characteristics of the human mind.

Thus it came as a considerable shock to the intellectual community of the early 19th century to learn of the discovery of non-Euclidean geometries, in which one of Euclid's axioms (the parallel postulate) did not hold. One significant consequence of this discovery was that philosophers and mathematicians began to regard axioms not as self-evident truths but rather as arbitrary rules, subject only to the requirement of consistency. For the axioms to be great the theory derived from them had to be both beautiful and useful, but to be mathematics the theory had only to be consistent.

Yet in spite of the widespread 20th-century consensus on consistency as the sole criterion of mathematical truth, the vast majority of mathematicians and scientists maintain a Platonic view of the continuum. Nearly everyone who studied calculus in the past 50 years was expected either to have a clear intuitive (Platonic) image of the real-number line or to believe that all its properties were consequences of some 10 to 15 supposedly self-evident truths that describe what is called formally a "complete ordered field" and informally the "real-

number system" [*see illustration on this page*].

A cynic might view this program as indoctrination in Platonism, since its clear purpose is to convince students of something their professors have accepted, namely that there is a unique real-number line with certain self-evident properties, and that mathematicians have succeeded, by listing the axioms of a complete ordered field, in capturing the essence of this line in a dozen or so sentences. Thus mathematicians now talk about *the* real-number line just as mathematicians and philosophers of the 18th century talked about *the* geometry.

Of course, the existence of different geometries does not necessitate the existence of different real-number lines. Nonetheless, in 1931 Kurt Gödel showed that in any mathematical system sufficiently large to contain arithmetic, there will always be undecidable sentences: statements about the system that can be neither proved nor disproved by logical deduction from the axioms [see "Gödel's Proof," by Ernest Nagel and James R. Newman; SCIENTIFIC AMERICAN, June, 1956]. Gödel's now famous "undecidability theorem" implies that in the Platonic universe of ideal mathematical objects there are many—in fact, infinitely many—objects that satisfy the axioms for the real-number line, since each undecidable proposition about the real-number line may be true in one ideal model and false in another.

In 1963, more than 30 years after Gödel proved that different models for the real-number axioms must exist, Paul J. Cohen of Stanford University actually constructed some models in which Georg Cantor's famous "continuum hypothesis" was false [see "Non-Cantorian Set Theory," by Paul J. Cohen and Reuben Hersh; SCIENTIFIC AMERICAN, December, 1967]. Cohen's methods have been applied extensively over the past eight years to yield a large variety of alternative models in all areas of mathematics. Later in this article I shall show how to construct one of these models for the real-number line, but first I should like to discuss an entirely different source of alternative models, since Gödel's undecidability theorem is not the only line of attack on the Platonic ideal of the real-number line.

Ever since Isaac Newton and Gottfried Leibniz laid the foundations of calculus in the late 17th century, mathematicians, philosophers and physicists have been quarreling about whether or not the real-number line contains infinitely small objects called infinitesimals. Infinitesimals played a key role in the development of the definitions and notations of calculus. Indeed, during the 18th and 19th centuries the name for what we now call simply "calculus" was "calculus of infinitesimals." What Newton and Leibniz did was to show how one could calculate with infinitesimals and obtain reliable results, results that moreover could not be obtained by any other method.

The calculus of infinitesimals was received with profound skepticism by many philosophers who, following Aristotle, abhorred the absolute infinite. In his *Physics* Aristotle distinguished between the potential infinite and the absolute infinite, accepting the former but rejecting the latter as untenable, or beyond the firm grasp of the human mind. (In taking this position Aristotle was merely reflecting the widespread Greek mistrust of the infinite, most popularly illustrated by the paradoxes of Zeno.) The philosopher Leibniz, somewhat chagrined at what the mathematician Leibniz had wrought, squirmed out of the dilemma by describing infinitesimals as "fictions, but useful fictions." Meanwhile Bishop Berkeley scorned Newton's infinitesimals (or fluxions, as Newton called them) as "ghosts of departed quantities."

But mathematics flourished in spite of the philosophers, awakened by the calculus of infinitesimals as it had not been since the glories of Athens and Alexandria. The Platonic image of the real-number line was as yet only a vague ideal, and calculus developed more as a descriptive science than as a deductive logical system. It was not until the 19th century that mathematicians began to echo the philosophers' skepticism; it was only then, as the axioms for the real-number system were distilled from the great unorganized mass of mystical properties, that it became clear that the existence of infinitesimals was inconsistent with these axioms.

This inconsistency is an immediate consequence of one of the most char-

1. Addition and Multiplication
If x and y are real numbers, then so are $x + y$ and xy.

2. Associativity
If w, x and y are real numbers, then $(w + x) + y = w + (x + y)$ and $(wx)y = w(xy)$.

3. Commutativity
If x and y are real numbers, then $x + y = y + x$ and $xy = yx$.

4. Distributivity
If w, x and y are real numbers, then $w(x + y) = wx + wy$.

5. Identities
There exist two special numbers z (or 0) and u (or 1) called the zero and the unit that satisfy $x + z = x$ and $xu = x$ for all real numbers x.

6. Additive Inverse
If x is any real number, there is another real number denoted by $-x$ and called the negative or additive inverse of x that satisfies $x + -x = z$, where z is the zero.

7. Multiplicative Inverse
If x is any real number except zero, there is another real number x^{-1}, called the reciprocal or multiplicative inverse of x, that satisfies $x\,x^{-1} = u$, where u is the unit.

8. Trichotomy
If x and y are real numbers, then either $x < y$ or $x = y$ or $x > y$.

9. Transitivity
If $w < x$ and $x < y$, then $w < y$.

10. Isotony
If $x < y$, then $x + w < y + w$; if $x < y$ and $w > z$ (where z is zero), then $xw < yw$.

11. Completion
Suppose E is a set of real numbers that has an upper bound, that is, suppose there is some real number x such that $y < x$ whenever y is a member of the set E. Then E has a least upper bound, that is, an upper bound x that is less than or equal to every other upper bound for E.

ELEVEN AXIOMS are customarily used to define the real-number line, which is known formally as a complete ordered field. The first seven axioms define a field; the first 10 define an ordered field. The symbol $>$ means "greater than"; the symbol $<$ means "less than."

acteristic properties of infinitesimals, namely that any multiple of an infinitesimal is still an infinitesimal. For instance, the infinitesimal dx used in calculus is smaller than every ordinary positive real number, and so is every multiple $m(dx)$ for any positive integer m. Accordingly the set M of all multiples of the infinitesimal dx has many upper bounds (any ordinary positive real number will do), but it has no least upper bound, since for any given number b that is an upper bound for M the smaller number $b - dx$ will also be an upper bound for M. Thus M fails to satisfy the final axiom for the real-number system, the "completeness" axiom.

Hence at the same time that geometers were being forced to change their criterion of truth from self-evidence to consistency, analysts were discovering that the supposedly self-evident axioms for the calculus of infinitesimals were inconsistent. What was to be done? The answer, developed principally by Augustin Cauchy and Karl Weierstrass, was to abandon the infinitesimals but keep the calculus (whence our present abbreviated name for the subject). Cauchy reformulated the foundations of calculus by substituting the concept of a limit for that of an infinitesimal; his method was a return to the Aristotelian concept of the potential infinite as the only secure basis for reasoning.

Weierstrass extended Cauchy's work by defining the concept of a limit in terms of the more primitive concept of real numbers. The Cauchy-Weierstrass approach to calculus, now widely taught, placed the epistemological foundation of calculus squarely on the shoulders of the real-number line. Although many users of calculus continue to prefer the intuitive language of infinitesimals, virtually all 20th-century mathematicians have adopted the definitions and concepts of Cauchy and Weierstrass.

Within the past decade, however, Abraham Robinson of Yale University developed a consistent mathematical theory of infinitesimals. This new theory, called "nonstandard analysis," resuscitated the discredited ideas of the actual infinite and actual infinitesimal and showed how much of modern mathematics could be consistently translated into a language of infinities and infinitesimals. In particular Robinson's new theory created a significant alternative to the mathematician's real-number line (alias the complete ordered field), an alternative that contained infinitesimals and on which calculus could be done in the spirit of Newton and Leibniz.

Thus by the end of the 1960's there were available on two different fronts several pretenders to what can justifiably be called the throne of mathematics: the real-number line. Although the technical construction of each of these alternatives is long and complex, there is a comparatively simple approach to both Cohen's and Robinson's models through the theory of probability; I shall now outline this approach, emphasizing its spirit more than its detail.

Before beginning this task it would be well to consider a somewhat subtle philosophical issue. If we were attempting to prove the real-number line defective and to replace it with a better model, we would have to take great care to avoid assuming the existence of the real-number line while constructing the alternative. This, however, is not what we are trying to do. We are, rather, trying to establish the existence of several different models of the real-number line, including the ordinary Platonic ideal of the complete ordered field. Therefore in constructing our new models we shall feel free to use the old one. (This process is quite analogous to the one followed in the construction of the non-Euclidean geometries, where the new models are defined within the standard Euclidean space.)

Let us now proceed to construct two specific models for the real-number system: one that will contain infinitesimal elements and another that will contain a set that violates Cantor's continuum hypothesis. Our method will be to construct a general model and then to obtain from it the two special models. The symbol $\boldsymbol{R}$ will be used throughout the discussion to stand for the ordinary real-number line; when we write $x \in \boldsymbol{R}$, we mean that x belongs to, or is a member of, the set $\boldsymbol{R}$, that is, x is a real number.

The objects in our new models will be real-valued functions f defined on some set S. In other words, f is a rule that assigns to each point s belonging to the set S a real number $f(s)$ belonging to the real-number set $\boldsymbol{R}$. The symbolism $f : S \to \boldsymbol{R}$ expresses the fact that f causes each point of S to be assigned a value in $\boldsymbol{R}$; the expression is usually read as "f maps S to $\boldsymbol{R}$" or, more formally, as "f is a function from S to $\boldsymbol{R}$." We shall use the symbol $\boldsymbol{R}^S$ to denote the set of all functions from S to $\boldsymbol{R}$.

In order for us to see how $\boldsymbol{R}^S$ can begin to resemble the real-number system, we must understand how to do addition and multiplication. Since the elements of $\boldsymbol{R}^S$ are functions (that is, rules), one must in effect define the addition and multiplication of rules. If f and g are functions in $\boldsymbol{R}^S$, we shall define $f + g$ to be the new function in $\boldsymbol{R}^S$ that assigns to each point s belonging to the set S the real number $f(s) + g(s)$; in pure symbols, $(f + g)(s) = f(s) + g(s)$. Multiplication is similar: $(fg)(s) = f(s)g(s)$. It is quite easy to show that the functions in $\boldsymbol{R}^S$ satisfy the first six axioms of a complete ordered field, listed in the illustration on the preceding page. A simple example, however, suffices to demonstrate that Axiom No. 7, the "multiplicative inverse" axiom, is not satisfied by this set [*see illustration at left*]. Thus $\boldsymbol{R}^S$ fails even to be a field and consequently falls far short of being a model for the real-number line.

To correct this situation we shall engage in a bit of mathematical gerrymandering. Axiom No. 7 does not say that all elements of $\boldsymbol{R}^S$ must have inverses; it only says that if it is true that f does not equal zero, then f must have an inverse. Accordingly if we have some functions f (that is, elements of $\boldsymbol{R}^S$) that do

1. Suppose S is the two-point set $\{0,1\}$.

2. Then each function $f: S \to \boldsymbol{R}$, which causes each point of S to be assigned a value in the real-number set $\boldsymbol{R}$ can be defined by giving two real numbers, namely the value of f at 0 and the value of f at 1.

3. The identity element u belonging to the trial set $\boldsymbol{R}^S$ is given by $u(0) = u(1) = 1$, and the zero element z belonging to $\boldsymbol{R}^S$ is given by $z(0) = z(1) = 0$.

4. Now the function h belonging to $\boldsymbol{R}^S$ defined by $h(0) = 0$ and $h(1) = 1$ fails to satisfy Axiom No. 7 since h is not equal to z and yet h does not have an inverse.

5. Therefore the set $\boldsymbol{R}^S$ is not an adequate model of the real-number line.

FAILURE OF $\boldsymbol{R}^S$ (the set of all functions from some arbitrary set S to the real-number set $\boldsymbol{R}$) to satisfy Axiom No. 7, the "multiplicative inverse" axiom, is demonstrated in this example. For the purpose of the demonstration S is assumed to be the two-point set $\{0,1\}$. The conclusion is that $\boldsymbol{R}^S$ falls far short of being a model for the real-number line.

not have inverses, we shall simply redefine truth so that for them the statement "f equals zero" will be true. More precisely, we shall substitute for the absolute notion of truth the more flexible concept of probable truth. In order to show how this can be carried out, I must digress for a while to discuss some definitions and examples from elementary probability.

Mathematical probability is based on a special function that assigns to each subset A of a given set Ω a positive real number that represents the probability that a point selected "at random" from the set Ω will actually be in A. This function is called a "probability measure" on the set Ω, and we shall denote it by m. The function m can be thought of as a rule that measures the size of sets, and the real number $m(A)$ can be thought of as the measure, or size, of the set A. Since the probability is 1 that a point selected at random from Ω will be in Ω, the measure of Ω must be 1. In addition we require of m only that it satisfy the (self-evident?) maxim that the whole is equal to the sum of its parts: if the whole set A is broken down into finitely or infinitely many distinct parts $A_1, A_2, \ldots, A_n, \ldots$, then $m(A) = m(A_1) + m(A_2) + \ldots + m(A_n) + \ldots$.

Before applying this measure to our set $\boldsymbol{R}^S$ I shall give two important examples of measures that will be used as the basis for our two different models for the real-number line. First, suppose Ω is the set N of positive integers: in symbols, $N = \{1, 2, 3, \ldots\}$. We shall define on N a very crude measure by classifying subsets of N as small if they are finite and large if they are cofinite, that is, if their complement (the set of positive integers not in them) is finite. Any small set will be given measure 0 and any large set will be given measure 1.

Now, the measure on N as just defined fails to measure sets such as the set of even integers, since neither it nor its complement is finite. The problem is that the set of even integers is neither small nor large but somehow in between. We can correct this deficiency in our measure m by systematically classifying each intermediate set as either small or large on an arbitrary basis, subject only to the constraints of consistency: each subset of a small set must be small; each set that contains a set previously classified as large must be large; the complement of a small set must be large, and vice versa. Once this has been done our measure will assign a value of either 0 or 1 to every subset of N; we shall call m the cofinite measure on N.

For our second example let us take for the set Ω the unit interval I that consists of all real numbers between 0 and 1; in symbols, $I = \{x \varepsilon \boldsymbol{R} \mid 0 \leqslant x \leqslant 1\}$. We build up a measure m on the subsets of I by first assigning to each interval its length. Then for a subset A that consists of pieces that are intervals we use the maxim that the whole is equal to the sum of its parts to say that the measure of A is the sum of the lengths of its pieces. Continuing in this fashion, we can build a probability measure on I that is known as the Lebesgue measure (named after the 20th-century French mathematician Henri Lebesgue).

Since each point x belonging to the set I is considered to be an interval of length 0, the Lebesgue measure m of each point x is 0. According to the construction just described, each finite set will also have measure 0, as will each infinite set that can be written as a sequence $x_1, x_2, x_3, \ldots$ (since the measure of the entire sequence is just the sum of the measures of each of its points). Since the set of rational numbers in I can be listed in a sequence (1, 1/2, 1/3, 2/3, 1/4, 2/4, 3/4, ...) it too must have measure 0, even though it appears to be a very large set. In other words, the chances are nil that a number selected at random from the unit interval will be rational.

This minor paradox leads directly to a major paradox concerning the Lebesgue measure. If I select a number at random from the unit interval, the probability is 0 that the number so chosen will equal some particular previously selected number; intuitively there are just too many numbers in the unit interval to choose from. Yet the probability that I shall pick some number in the unit interval is 1. Thus in this case it appears that the whole is not equal to the sum of its parts! This dilemma has plagued various philosophers throughout history; how can something of positive weight be made up of parts that each have zero weight?

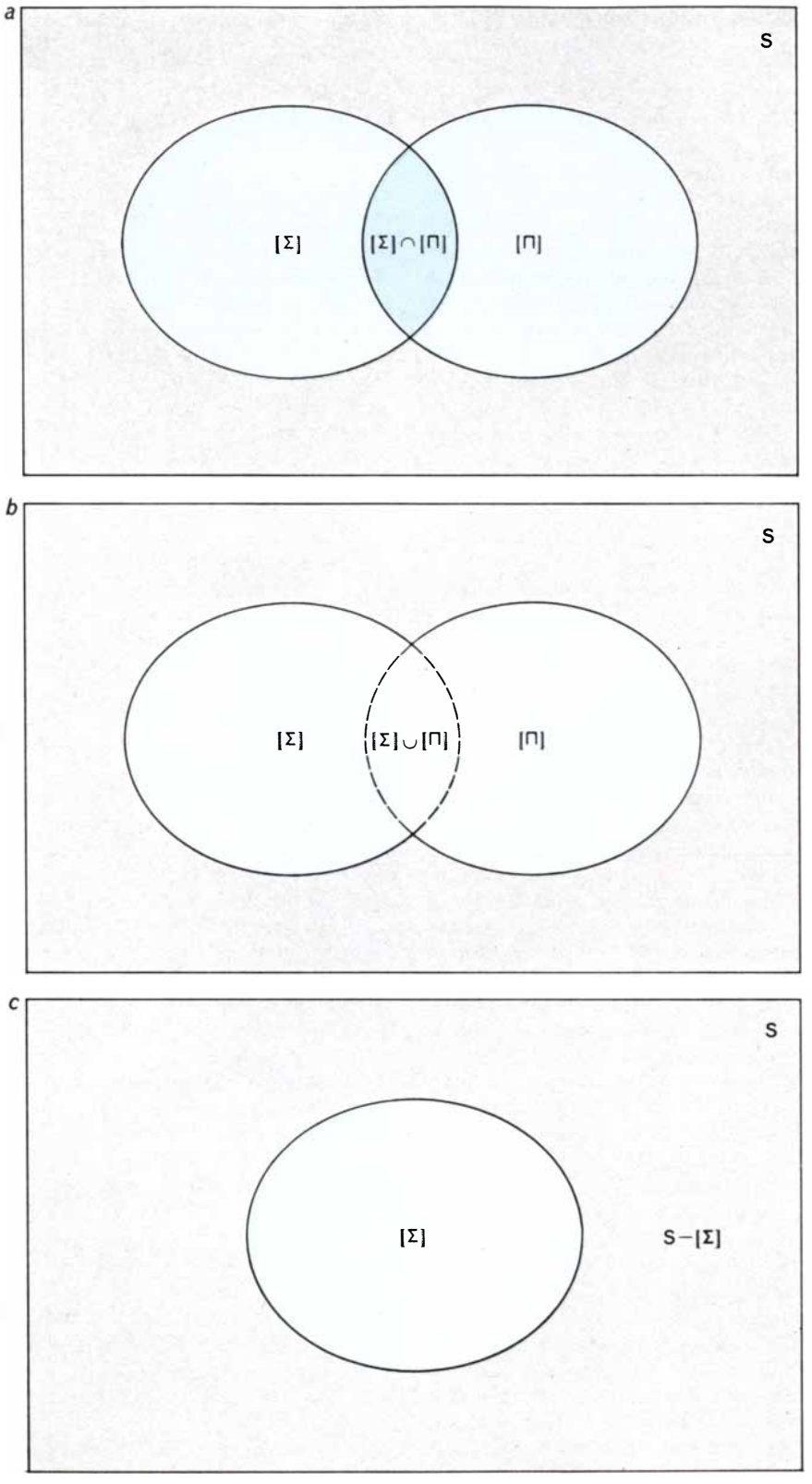
The trouble is that there are too many points in the unit interval. Cantor called countable those infinite sets that can be written as a sequence $x_1, x_2, x_3, \ldots$, one point for each integer; other infinite sets are called uncountable. The principle that the whole is equal to the sum of its parts applies only if the collection of parts is finite or countable. Hence the measure of the set of rational numbers given above is 0 because the set of rationals can be written as a sequence in which each term has measure 0. Similar reasoning does not allow us to conclude that the measure of the interval I is 0 because I cannot be written as a sequence. (The fact that the unit interval I is an uncountable set is actually a very famous theorem of Cantor's.) It follows that $\boldsymbol{R}$, which contains I, is also uncountable.

We now return to our original task of building models for the real-number line. We had constructed the set $\boldsymbol{R}^S$ of functions from S to $\boldsymbol{R}$ and showed how this set failed to satisfy Axiom No. 7 for complete ordered fields. In order to correct this defect in $\boldsymbol{R}^S$ we shall now assume that S is equipped with a probability measure m, so that we shall be able to measure the size of subsets of S.

The most primitive statement that can be made concerning two functions in $\boldsymbol{R}^S$ is that they are equal (or unequal). If f and g are given, then "f equals g" is either true or false. It is true if and only if $f(x)$ equals $g(x)$ for all points x belonging to the real-number set $\boldsymbol{R}$. In other words, f equals g if and only if f and g express the same rule. If there is in $\boldsymbol{R}$ so much as one point y where $f(y)$ does not equal $g(y)$, then we are forced to say that the function f does not equal the function g—even if they agree at every other point of $\boldsymbol{R}$. Logicians call the terms "true" and "false" truth values. In traditional logic (which we have used throughout this discussion) sentences have only two truth values: true or false.

We propose to change this tradition by assigning to sentences not just one of two possible truth values but a real number that expresses the probability of truth; this number may be 0 (if the sentence has no chance of being true) or 1 (if the sentence is certain to be true) or any number in between. We can accomplish this *coup d'état* on $\boldsymbol{R}^S$ because we now have a means of measuring the size or probability of subsets of S. In particular we shall say that the probability that "f equals g" is true is the measure of the set of points in S where it really is true, that is, the measure of the set of all points s belonging to S that satisfy the requirement that $f(s)$ equal $g(s)$. If we denote this probability truth value of "f equals g" by $|f = g|$, we have $|f = g| = m(\{s \varepsilon S \mid f(s) = g(s)\})$.

Of course, there are a lot of sentences about $\boldsymbol{R}^S$ that are far more complicated than the primitive sentences of the form "f equals g." All the more complex sentences, however, can be built up from a few primitive ones, which are called, naturally enough, atomic sentences. What we have to do to make our revolution in truth values succeed is, beginning with the atomic sentences, to systematically work our way through the



CONSIDERABLE SIMILARITY exists between the rules that govern the use in logic of the connectives "and," "or" and "not," and the operations on sets of, respectively, intersection (*a*), union (*b*) and complementation (*c*). This similarity (in fact, a mathematical isomorphism) was discovered by the 19th-century English mathematician George Boole, and the abstract system based on it is now called Boolean algebra. The symbol $[\![\Sigma]\!]$ stands for the set where the sentence Σ is true; $[\![\Pi]\!]$ signifies the set where the sentence Π is true. The intersection of these two sets, symbolized $[\![\Sigma]\!] \cap [\![\Pi]\!]$, is the set where the sentence "Σ and Π" is true; the union of these two sets, $[\![\Sigma]\!] \cup [\![\Pi]\!]$, is the set where "$\Sigma$ or Π" is true. In the example of complementation $S - [\![\Sigma]\!]$ signifies the set where "not Σ" is true.

entire catalogue of sentences about $\boldsymbol{R}^s$ and determine how to assign a probability to each sentence in such a way that it represents the measure of the set of points in S where the sentence is true.

The basic idea behind this scheme was outlined more than 100 years ago by the English mathematician George Boole, who noticed that there was a considerable similarity (in fact, a mathematical isomorphism) between the rules that govern the use in logic of the connectives "and," "or" and "not," and the operations on sets of, respectively, intersection, union and complementation [*see illustration at left*]. The abstract mathematical system that represents this structure is now called a Boolean algebra. To see how Boole's system can be used to solve our problem let us consider a simple example.

If Σ is a sentence about $\boldsymbol{R}^s$, we shall denote by $[\![\Sigma]\!]$ the set of all points in S where Σ is true, and by $|\Sigma|$ the measure of $[\![\Sigma]\!]$. Thus $|\Sigma| = m([\![\Sigma]\!])$ is the probability truth value of the sentence Σ. If Π is another sentence about $\boldsymbol{R}^s$, then one of the Boolean relationships is that $[\![\Sigma \text{ and } \Pi]\!] = [\![\Sigma]\!] \cap [\![\Pi]\!]$. In words, the set of points in S where both Σ and Π are true is the intersection (denoted by $\cap$) of the set where Σ is true with the set where Π is true.

Perhaps the only detail in our plan (to extend the probability truth values from atomic sentences to all sentences) that requires special comment is the treatment of the quantifiers $\forall$ (meaning "for all") and $\exists$ ("there exists"). The sentence "$\forall s \Pi(s)$," which we read as "For all points s belonging to the set S, the sentence $\Pi(s)$ is true," means that $\Pi(s_1)$ and $\Pi(s_2)$ and ... (ad infinitum) are all true. In other words, $\forall$ is an infinite repetition of "and." Thus it is appropriate that the set operation that corresponds to $\forall$ is the infinite intersection, since the ordinary (finite) intersection corresponds to "and." Similarly, $\exists$ represents an infinite "or," so that its corresponding set operation is the infinite union.

In summary, then, although we have by no means discussed all the relevant detail, it is possible by means of the Boolean translation to determine for each sentence Σ about $\boldsymbol{R}^s$ a subset $[\![\Sigma]\!]$ of S on which Σ is true. The measure of this set is the probability truth value of Σ, denoted by $|\Sigma|$. A sentence Σ will be considered valid if and only if $|\Sigma| = 1$; in other words, Σ is valid if and only if the measure of the set where Σ fails to hold is 0, so that there is "no chance" of failure. The examples of measures given

above reveal the essential distinction between the concept of truth in the two-valued logic and the concept of validity in the probability-valued logic; for Σ to be true it must hold without exception for all points s belonging to the set S, but it remains valid even if it fails to hold on a fairly large subset, provided only that the measure of the exceptional set is 0.

Let us denote by $\boldsymbol{R}^S/m$ the set $\boldsymbol{R}^S$ with the new concept of validity as determined by the measure m. We have seen that under the former notion of truth $\boldsymbol{R}^S$ did not satisfy the axioms for a complete ordered field. But $\boldsymbol{R}^S/m$ will always be an ordered field because, on the basis of the concept of validity, it will always satisfy axioms No. 1 through No. 10. Furthermore, for certain choices of S and m it will also satisfy the completion axiom, so that in these cases $\boldsymbol{R}^S/m$ will be a complete ordered field.

Suppose S is the set of positive integers N with the cofinite measure m defined above in our first example of measures; m assigns to finite sets the value 0, to cofinite sets the value 1 and to intermediate sets either 0 or 1 according to some arbitrary but consistent pattern. In this case we can picture each function f belonging to $\boldsymbol{R}^S$ as a sequence $f_1, f_2, f_3, \ldots$, where for each integer i belonging to $N = S$, f_i is the real number that f assigns to i; f_i is just what we usually call $f(i)$. For example, the identity function u is the sequence $1, 1, 1, 1, \ldots$, since $u_i = u(i) = 1$ for each i belonging to $N = S$.

Given S and m as defined above, $\boldsymbol{R}^S/m$ contains both infinitesimals and infinite elements. The function f defined by the sequence $1, 1/2, 1/3, 1/4, \ldots$ is an infinitesimal, since if we multiply it by any integer n, the resulting sequence $n, n/2, n/3, n/4, \ldots$ is smaller than the identity element u except for the finite set $\{n, n/2, n/3, \ldots, n/(n-1), n/n\}$, which has measure 0. In other words, if $\Pi(n)$ is the sentence "nf is less than u," then $|\Pi(n)| = 1$ because $\{i \in N | nf(i) < u(i)\}$ has measure 1. Thus $|\forall n \Pi(n)| = 1$, which is to say that for every n, nf is less than u. Thus f is an infinitesimal since only an infinitesimal could have this property.

For a different perspective let us look at the multiplicative inverse of f; f must have an inverse since $\boldsymbol{R}^S/m$ is a field. In reality the function g defined by the sequence $1, 2, 3, 4, \ldots$ is the inverse of f since, clearly, $gf = u$. Just as f is an infinitesimal, so g must be infinite. In fact, if n denotes the constant function $n, n, n, \ldots$, then $|g > n| = m(\{i | g_i > n_i\}) = m(\{i | i > n\}) = 1$, since the set in question is cofinite. Thus in $\boldsymbol{R}^S/m$ it is valid to say that g is greater than n for all n. This is just what we mean when we say that g is infinite.

We have now achieved the first of our two goals: the description of a mathematically consistent model of the real numbers that contains infinitesimals. The other type of model, in which Cantor's continuum hypothesis fails, requires some more work. First, however, I shall explain Cantor's hypothesis.

Cantor developed his theory of cardinal numbers by defining two sets as being of equal size (or of the same "cardinality") if there is some function, or rule, that establishes a one-to-one correspondence between them. All finite sets with the same number of elements have the same cardinality, and all infinite sequences have the same cardinality as the set of integers. (The sets in this latter category are those that we have called countable.) Cantor's great achievement was to assign sizes (cardinal numbers) even to the uncountable sets. In particular he identified a class of sets of the same cardinality as the set of real numbers $\boldsymbol{R}$; he called the cardinal number of sets in this class c, for continuum. Sets of size c are those that can be put into one-to-one correspondence with the real numbers.

Cantor's continuum hypothesis is simply this: Every infinite set of real numbers is either countable or of cardinality c. There are no sets of intermediate size. This hypothesis was formulated (but not proved) by Cantor late in the 19th century, and it was listed in 1900 by David Hilbert as the first of his celebrated 23 problems for 20th-century mathematics. We shall now outline a version of Cohen's 1963 model that shows conclusively that Cantor's continuum hypothesis could never be proved from the ordinary axioms of set theory and real numbers.

This new model, developed principally by Dana S. Scott of Princeton University, involves the use of a rather complicated set S in the general model $\boldsymbol{R}^S/m$. We first have to find a very large set T, a set whose cardinality is bigger than c, the cardinal number of $\boldsymbol{R}$. The proof that such sets exist is another of Cantor's significant contributions to mathematics. An example of such a set is what is called the power set of $\boldsymbol{R}$: the set that consists of all subsets of $\boldsymbol{R}$. Cantor's proof that the cardinal number of this set is indeed bigger than c is essentially the same as his proof that $\boldsymbol{R}$, or equivalently the unit interval I, is strictly bigger than the set of integers.

Let I denote the unit interval ($I = \{x \in \boldsymbol{R} \mid 0 \leqslant x \leqslant 1\}$), and let T be a set whose cardinal number is bigger than c. Then we define S to be the set $I^T = \{f : T \to I\}$. The points of S are functions from T to I. The measure on S, whose details need not concern us, is a rather natural extension to I^T of the Lebesgue measure on I described above. Our new model for the real-number line is $\boldsymbol{R}^S/m$, where S equals I^T.

The fact that the points of S are really functions can be the source of some conceptual difficulty if we confuse the functions in $S = I^T$ with those in $\boldsymbol{R}^S$. Mathematicians avoid this confusion by emphasizing in their minds that S is a *set* of functions rather than a set of *functions*. We put up with this double meaning because the entire key to Scott's model is a deliberate pun formed by interchanging the role of functions and points.

To be precise, we observe that each point t belonging to the set T can be thought of as a function in $\boldsymbol{R}^S$ that assigns to each function f belonging to $I^T = S$ the real number $f(t)$ belonging to I. If we call the function t by the name $\hat{t}$, then $\hat{t}(f)$ equals $f(t)$ whenever f is a member of S. Now our problem is to find a subset of $\boldsymbol{R}^S$ that is uncountable, yet not as large as $\boldsymbol{R}^S$. But just how large is $\boldsymbol{R}^S$? Since it contains a $\hat{t}$ for each t belonging to T, the cardinality of $\boldsymbol{R}^S$ must be at least as large as that of T, but T was selected so that its cardinality was larger than c. Hence the cardinality of $\boldsymbol{R}^S$, our new real-number line, is greater than c, the cardinality of the old real-number line $\boldsymbol{R}$.

Since the new real-number line is so very much bigger than the old one, it will be much easier to find in the new line an uncountable set that is not of the same cardinality as the entire line. All we have to do is select a subset $\hat{P}$ of T with the property that the cardinality of P is c. (If we had been thinking of T as the power set of $\boldsymbol{R}$, then we could take $\boldsymbol{R}$ itself for P.) Then the set P that consists of all the functions $\hat{t}$ for t belonging to P is a subset of $\boldsymbol{R}^S$ of size c. Thus $\hat{P}$ is uncountable, yet it is not as large as $\boldsymbol{R}^S$.

This completes our program of constructing new models of the real-number line. To summarize briefly, each model consists of functions from some set S to the ordinary real-number line $\boldsymbol{R}$. The two-valued true-false logic of the old system is replaced with a more flexible probability-valued logic determined by some measure on the set S. Validity in the new model means truth with probability 1, a definition that grants exceptions to sets of measure 0. The particular models that contain infinitesimals or

counterexamples to the continuum hypothesis are then formed by selecting special sets S with special probability measures m.

In the preceding development of the two new models of the real-number line I not only omitted all proofs and technical detail but I also failed to mention a most important distinction between these two models: the Cohen-Scott model (where S equals I^T) is a complete ordered field, whereas the Robinson model (where S equals N) is just an ordered field. The fact that Robinson's model fails to be complete should come as no surprise, for we did see that any real-number line that contained infinitesimals could not satisfy Axiom No. 11. To complete the discussion of these models I should explain two things: Why should the two models behave differently with respect to Axiom No. 11, and by what right can we call an object (such as Robinson's model) that fails to satisfy Axiom No. 11 a real-number line?

The answers to both questions depend on a distinction that logicians make between what are called first-order sentences and higher-order sentences. Intuitively a first-order sentence about the real numbers is one that makes generalizations only about real numbers and not about such higher abstractions as functions or sets of real numbers; in technical terms a first-order sentence is one in which the quantifiers ($\forall$ and $\exists$) are limited to real numbers. All the axioms of a complete ordered field are first-order sentences except for the completion axiom (No. 11), which talks about a property of *all* subsets.

Now, the models of nonstandard analysis (such as the first of our two examples) have the property that every first-order sentence that is true about $\boldsymbol{R}$ is true about them, and conversely. In other words, it is impossible to distinguish between these new models and $\boldsymbol{R}$ by the use of first-order sentences. Since a large part of calculus can be expressed in first-order sentences, it is quite literally true that for most practical purposes the nonstandard real-number line is the same as the ordinary real-number line.

The effort required to show that the completeness axiom is satisfied by some particular model $\boldsymbol{R}^S/m$ is decidedly more strenuous than any of the ideas discussed above. Not only is the axiom itself more difficult to verify, but also before we can even consider it we must work quite hard to extend the machinery of assigning probability truth values from first-order sentences to higher-order sentences. This extra effort is also required to prove that in our second example the probability is zero that the continuum hypothesis is true—since the continuum hypothesis is also a higher-order sentence.

There is a striking analogy between the creation of these new models for the real-number line and the development of quantum mechanics. The most common derivation of quantum mechanics from classical mechanics is accomplished by substituting in the classical theory a wave packet or probability cloud for each classical particle. In the quantum model the location of the particle in the cloud is given by a probability distribution and the particle itself becomes a random variable.

The new models for the real-number line that were described above are formed in just this way. We substituted for the ordinary real numbers certain real-valued functions defined on a probability space. These functions are precisely what statisticians and probability theorists call random variables. The substitution of random variables for real numbers was accompanied by a corresponding change from a two-valued logic to a more general scheme in which the truth value of a statement about the real numbers is expressed by a probability. This transition from two-valued to probability-valued logic also occurs in the derivation of quantum mechanics from classical mechanics, since questions about the state of the mechanical system that in the classical framework could be answered by "Yes" or "No" can in quantum theory only be answered with a probability.

One of the best-known consequences of quantum mechanics is Werner Heisenberg's uncertainty principle, which places an absolute limit on the amount of information that can be obtained by a physicist through the process of observing a particle. Heisenberg's principle is derived by a process of self-reflection in which the observer analyses the effect that his actions have on the system he is observing.

In mathematics, analogously, one deduces from the action of writing down axioms and theorems (which is to the mathematician what experimental observation is to the physicist) a limitation on the amount of information that can be derived from a set of axioms (in the Gödel-Cohen case) and a limitation on the capability of axiom systems to describe models uniquely (in Robinson's theory). The first of these two limitations is the substance of Gödel's undecidability theorem, whereas the second is one version of a more general result known in mathematical logic as the Lowenheim-Skolem theorem. In different ways each limitation expresses a basic uncertainty in mathematics.

It is not surprising that the introduction of "uncertainty" results in mathematics was accompanied by fundamental

$\boldsymbol{R}$ = Real-number line

ϵ = "Belongs to"; "is a member of" (as in $x \epsilon \boldsymbol{R}$, that is, x belongs to, or is a member of, the set of real numbers $\boldsymbol{R}$; hence x is a real number)

$f: S \rightarrow \boldsymbol{R}$ = A function f that causes each point of the set S to be assigned a value in the real-number set $\boldsymbol{R}$

$\boldsymbol{R}^S$ = The set of all functions from S to $\boldsymbol{R}$

$m(A)$ = The probability measure of the set A

Σ = A sentence about $\boldsymbol{R}^S$

$[\![\Sigma]\!]$ = The set of all points in S where Σ is true

$|\Sigma|$ = The truth value of the sentence Σ, that is, the probability measure of $[\![\Sigma]\!]$; also written $m([\![\Sigma]\!])$

Π = Another sentence about $\boldsymbol{R}^S$

$\cap$ = Intersection of two sets

$\cup$ = Union of two sets

$\forall$ = "For all"

$\exists$ = "There exists"

c = Cantor's symbol for the cardinal number of the continuum

GLOSSARY OF SYMBOLS used in this article is provided in this illustration. The symbol c, used by Georg Cantor to signify continuum, denotes the cardinal number of sets in the class of infinite, uncountable sets of the same cardinality, or size, as the set of real numbers $\boldsymbol{R}$.

disagreement about their significance, just as Heisenberg's uncertainty principle precipitated among physicists a major dispute about its ultimate significance. Most physicists tend to agree with Niels Bohr that the uncertainty in quantum mechanics is a fundamental law of nature. Some, however, support the position of Albert Einstein, who argued that the uncertainty principle merely expresses a limitation on the present conceptual formulation of physics.

These two opposite poles in physics reflect approximately the positions of the Formalist and the Platonist in mathematics. The Formalist believes that mathematics is pure form, and since the "uncertainty" theorems of mathematics limit the scope and power of the formalism with which mathematics is concerned, then they must forever limit mathematics itself. For a Formalist the question of whether the "real" real-number line satisfies the continuum hypothesis or contains infinitesimals is as meaningless as is, for a physicist such as Bohr, the question of whether or not an electron "really" has an exact simultaneous position and velocity. Robinson himself reflects this position when he writes (in a paper titled "Formalism 64") that "any mention of infinite totalities is literally meaningless."

In contrast, the Platonists, who count among their number even Gödel himself, believe (like Einstein) that the undecidability in mathematics is a statement about the inherent limitations of our present axiomatic mode of investigation and not about the mathematical objects themselves. Gödel has argued that there is no reason to believe more in the objective existence of a physical object such as an electron than in the objective existence of a mathematical object such as the real-number line. And one who believes that the real-number line has an existence in a Platonic universe can hardly avoid wondering about which of our several models is the more accurate description of this Platonic continuum.

It is ironic that Robinson, the re-creator of infinitesimals, does not believe they really exist, whereas Gödel, the prophet of undecidability, believes in a Platonic universe in which the properties of mathematical objects are visible for those who have the eyes to see. Perhaps these men are merely reflecting an intense modesty about the significance of their own achievements. It seems unlikely, however, that within the next few generations mathematicians will be able to agree on whether every mathematical statement that is true is also knowable.

Some things are changing for the better...

Whatever your job, here's a calculator that speaks your language. You can customize its keyboard, memory size, display, programs and peripherals to suit your number-crunching tasks.

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. This is the rationale behind every new instrument, computer or system that we tell you about in these ads. This month:

For picky people with particular problems: A design-your-own calculator

A user in virtually any discipline now can customize a powerful new programmable calculator to his specific computational needs.

An agronomist, for example, may want to examine the characteristics of a large plant population and determine the mean, standard deviation and standard error of their distribution. With the Model 10, he simply enters the raw data and hits a single key for the complete statistical analysis. A chromatographer can obtain per cent concentration and relative retention time of each component on his chromatogram . . . at a single keystroke. A physicist completes a sequence of acceleration, velocity, force and work . . . and a clinical pathologist computes a full blood gas analysis . . . at a single keystroke. Et cetera.

This is possible because the new Model 10 calculator has interchangeable function blocks which can define its keyboard to meet varying needs. One standard plug-in block emphasizes powerful statistical computations, another gives higher mathematics capability, and the third is completely user-definable. This block provides single keystroke solutions to multiple-step calculations commonly encountered by the user. Once programmed each key performs its customized function whenever he strikes it.

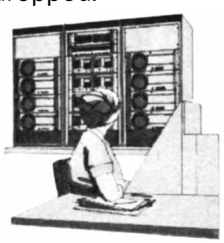
For more on tailoring the $2,975 Model 10 to your particular profession (full alphanumeric printing capability, expandable memory, a wide line of peripherals, etc.) write for our brochure.

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.

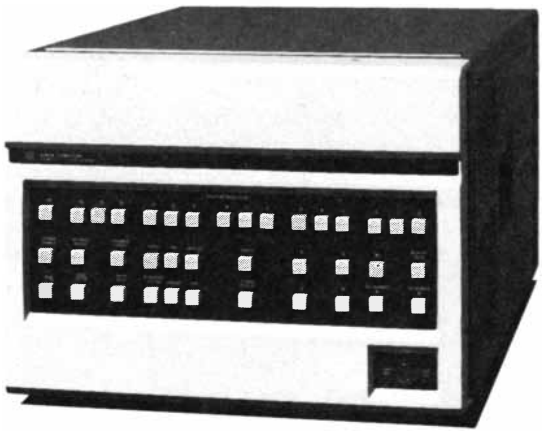


At the nursing station, the patient's ECG signal is monitored by a receiver that operates automatically, never requires tuning and accepts only valid signals, minimizing artifacts from patient motion. An automatic warning light alerts the nurse of such inoperative conditions as: patient out of range, 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.

Want to wield more power? Use a computer that has connections

Adding the power of a computer to the laboratory can change the role of a technologist for the better. That's especially true with HP's new 2100, a small computer that grows with your needs.

Once it rolls through your door, the 2100 resists replacement by expanding as your needs increase. Starting with its smallest configuration for only $6,900, you get a powerful computer that understands your instruments. Fourteen input/output devices and peripherals, and dozens of instruments, can be plugged directly into the 2100. HP has a big advantage here because we probably made many of your instruments in the first place.



The 2100 is more than just a pretty face. But if work, rather than beauty, turns you on, our box is simply gorgeous. Just imagine it plugged into the important tools of your trade—space doesn't permit showing them.

When you need more capability, you merely plug in additional memory and hook on peripherals. Smaller and nearly twice as fast as preceding models, the 2100 can expand from 4K to 32K of core memory, all within its 12-inch high mainframe. The same 2100 you start with can be upgraded to a time-sharing, batch processing or automated measurement, test or analysis systems.

The 2100 also gives you a wider choice of operating software packages than any other small computer. One other point: our 2100, as its predecessors, brings HP "instrumentation reliability" to computers. More information is yours for the asking.

For more complete information, write Hewlett-Packard, 1501 Page Mill Road, Palo Alto, California 94304. In Europe: 1217 Meyrin-Geneva, Switzerland. 00113

HEWLETT hp PACKARD

MATHEMATICAL GAMES

Ticktacktoe and its complications, and answers to the quickie puzzles

by Martin Gardner

"It's as simple as tit-tat-toe, three-in-a-row, and as easy as playing hooky. I should hope *we can find a way that's a little more complicated than* that, *Huck Finn."*

—Mark Twain,
The Adventures of Huckleberry Finn

Ticktacktoe (the spelling varies widely) is not nearly so simple as Tom Sawyer thought. When Charles Sanders Peirce wrote his *Elements of Mathematics,* a textbook that was never published, he included a 17-page analysis of this ancient game. It was one of Peirce's many anticipations of "modern math." Today's progressive teachers frequently use ticktacktoe to introduce their pupils to such concepts as the intersection of sets, rotational and mirror-reflection symmetry and higher Euclidean space. This month we consider some unusual aspects of the game not covered in a 1957 column (reprinted in *The Scientific American Book of Mathematical Puzzles & Diversions*) or in this department for February, 1967.

The traditional game, as most readers surely know, is a draw if both players do their best. From time to time pictures of a ticktacktoe game appear in advertisements and cartoons, and sometimes they provide pleasant little puzzles. For example, on May 13, 1956, in the New York *Herald Tribune,* there was an IBM advertisement with the unfinished game at the left in the illustration below. Which player went first, assuming that the players were not stupid? It takes only a moment to realize that *O* could not have gone first or *X* would have played the top left corner on his second move. The other two patterns are almost as trivial. Does the center one, from a Little Lulu cartoon in *The Saturday Evening Post* (January 16, 1937), depict a possible game? At the right below is a pattern from an advertisement by publisher Lyle Stuart in *The New York Times* (June 1, 1971). In which cell must the last move have been made?

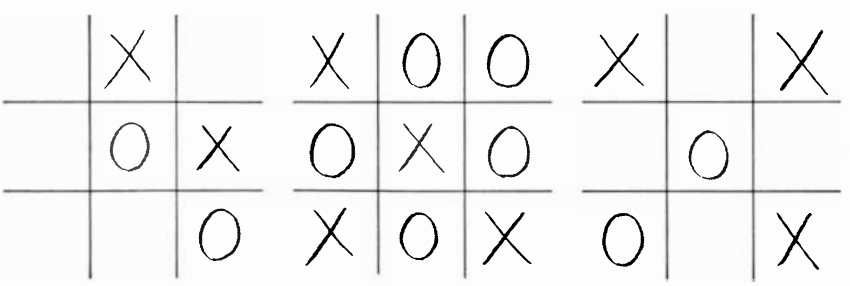
If the first player, say *X*, opens in the center cell, he can force a draw that always ends with the same basic pattern. This underlies several prediction tricks. For example, the magician draws the finish of a game, with all cells filled, on a square sheet of paper that he turns face down without letting anyone see it. He then plays a ticktacktoe game with someone, writing on another square sheet. After the game ends in a draw he turns over his "prediction." The two patterns match cell for cell.

The technique is explained in the top illustration on the opposite page. *X* plays the center opening. If *O* marks any corner cell, *X* forces the draw shown at the left in the illustration (moves are numbered in order of play). It is only necessary for *X* to remember where to make his second move, since all moves are forced from then on; a simple rule for the second move is to consider the corner opposite *O*'s first move and then play adjacent to it on the clockwise side. If *O* responds to the opening with a side cell, *X* forces the draw shown at the right. In this situation only *O*'s moves are forced and *X* must remember how to play his next four moves. The following simple rule was proposed by a magician who signed himself "Thorson" when he described this trick in the September 1960 issue of *M.U.M.*, official organ of the Society of American Magicians: *X* makes his second, third and fourth moves adjacent and clockwise to *O*'s previous moves, and his fifth move in the only remaining empty cell.

Note that the two final results are identical. Of course, each game can be played in any of four different orientations. The magician, recalling which corner of his inverted prediction has the *O* surrounded by three *X*'s, casually turns over the square sheet along the proper axis—orthogonal or diagonal—so that his predicted pattern matches the orientation of the game just completed.

The trick can even be repeated. This time *X* substitutes counterclockwise for clockwise in the rules, having drawn a prediction that is a mirror image of the preceding one. The two predictions will not match in any orientation and few people will realize that they are mirror reflections of each other.

Dozens of variations of planar ticktacktoe have been analyzed. Standard games on squares of higher order than 3, when the goal on an order-*n* board is to get *n* in a row, are uninteresting because the second player can easily force a draw. My 1957 column discussed games in which counters are moved over the board (one such version goes back to ancient Greece), and toetacktick, in which the first to get three in a row *loses.* A. K. Austin's "wild ticktacktoe," in which players may use either *X* or *O* on every move, was shown to be a first-player win in this department for July, 1964. What about "wild toetacktick," in which players can choose either mark on each move and the first three-in-a-row loses? In 1964 Solomon W. Golomb and Robert Abbott independently found that the simple symmetry strategy by which the first player can force at least a draw in standard toetacktick also applies to the wild version. A center opening is followed by playing directly opposite the other player's moves, always choosing *X* if he played *O* and *O* if he played *X*. The question remains: Does the first player have a *winning* strategy in wild toetacktick? Abbott made an exhaustive tree diagram of all possible plays and proved that the second player too can force a draw. Tame toetacktick



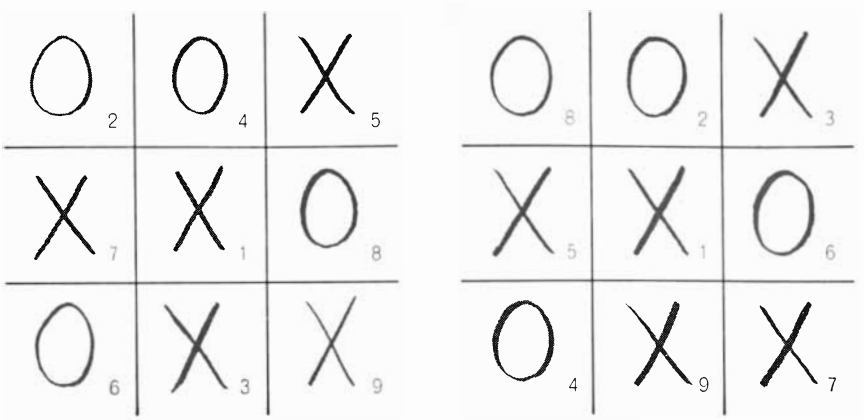
Three easy ticktacktoe puzzles

also is a draw if both sides play rationally.

An amusing variation appears in David L. Silverman's excellent new book of game puzzles, *Your Move*. The rules are the same as in standard ticktacktoe except that one player tries to achieve a draw and the other player wins if either of them gets three in a row. Can the reader show that no matter who plays first the player trying to force a row of three can always do so? Silverman does not answer this in his book, but I shall give his solution next month.

It is impossible to describe all the other planar variants that have been proposed, such as using numbers or letters as marks for the goal of forming a certain sum or spelling a certain word; playing on the vertexes of curious nine-point graphs (for a game on one such graph see this department for November, 1967); using counters with *X* on one side and *O* on the other, with the counters turned over according to specified rules. Games have been marketed in which flip-overs are randomized by concealed magnets that may or may not reverse a counter or by tossing beanbags at a board to cause cubical cells to alter their top symbols by rotating.

If ticktacktoe is played on an unlimited checkerboard, it is a trivial win for the first player if the goal is to get exactly four or any smaller number of one's marks in an orthogonal or diagonal row. When the goal is exactly five in a row (more than five does not count), this game is far from trivial. It is the ancient Oriental game known as go-moku (five stones) in Japan, where it is played on the intersections of a go board. (The game is sold in the U.S. by Parker Brothers under the name of Pegity.) Although it is widely believed that a first-player winning strategy exists, this has not yet, to my knowledge, been proved. (Geoffrey Mott-Smith is incorrect in claiming such a proof in his book *Mathematical Puzzles,* Dover, 1954.) Computer learning programs for go-moku have been devised (for one such program see *Machine Intelligence 2*, edited by E. Dale and D. Michie, Oliver & Boyd, 1968). In Japan the first player's strong advantage is weakened and the game prolonged by not allowing either player to establish two intersecting rows of three, each row open at both ends. Such a "fork" usually leads to an open-end row of four on the player's next move and a sure win on his following move. However, John L. Selfridge, a mathematician at the University of Illinois, has shown that if there is a first-player winning strategy without the fork prohibition, there must be one *with* the prohibition also.

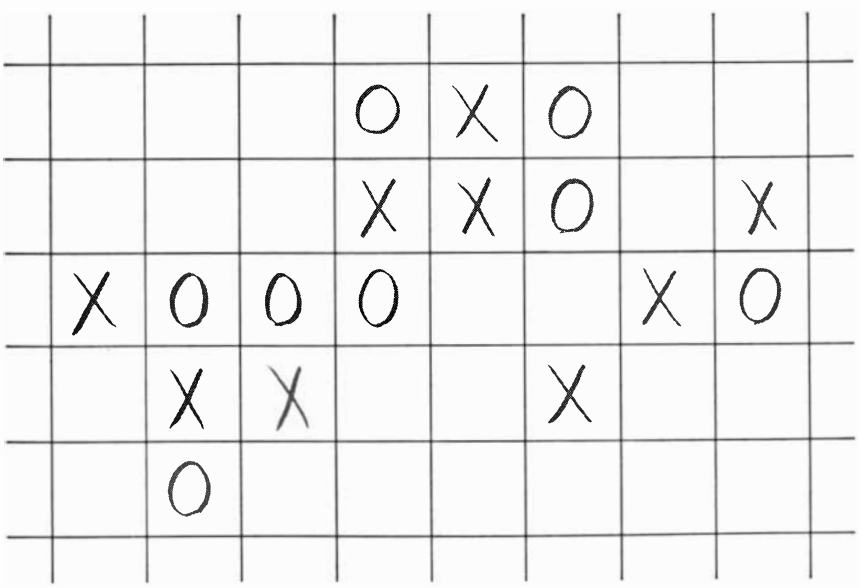


A ticktacktoe prediction trick

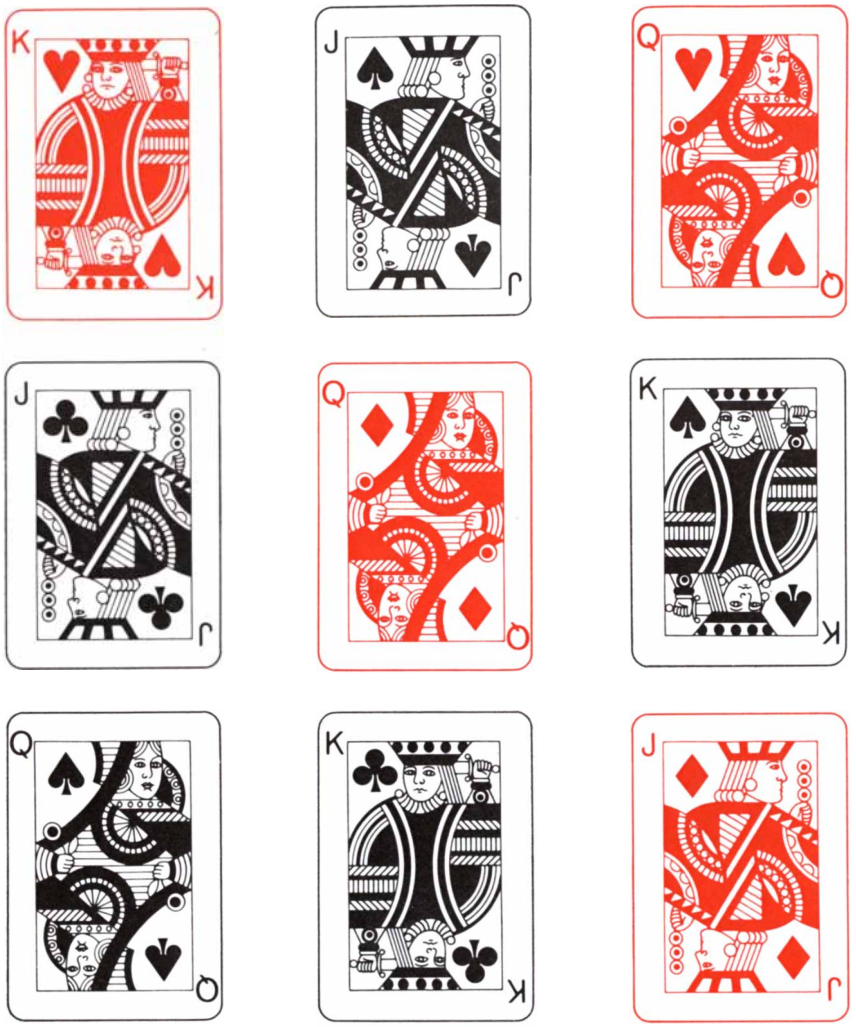
It is not possible that the second player has a winning strategy in go-moku or similar games in any dimension. The bare bones of the simple *reductio ad absurdum* proof, first formulated by John Nash for the game of hex, are as follows. Assume that a second-player winning strategy exists. If it does, the first player can make an irrelevant, random first play—a play that can only be an asset—and then, since he is now in effect the second player, win by appropriating the second player's strategy. Because this contradicts the assumption, it follows that no second-player winning strategy exists. The first player can either win or at least force a draw if the game allows draws.

Go-moku is a stimulating game. To catch its special flavor the reader is urged to study a position from Silverman's book [*see illustration below*] and determine how *O* can play and win in five moves. The answer will appear next month. Note that *X* has an open-end diagonal of three, which he threatens to lengthen to an open-end row of four.

When ticktacktoe is extended to three dimensions, the first player wins easily on an order-3 cube by first taking the center cell. As Silverman points out, if the first player fails to open with the center cell, the second player can win by taking it; if the center is permanently prohibited to both players, the first player has an easy win. Three-dimensional toetacktick (the first row of three loses) is also a win for the first player. He plays the same strategy used for forcing a draw in planar toetacktick: he first takes the center and then always plays symmetrically opposite his opponent. Since drawn positions are impossible on the order-3 cube, this technique forces the



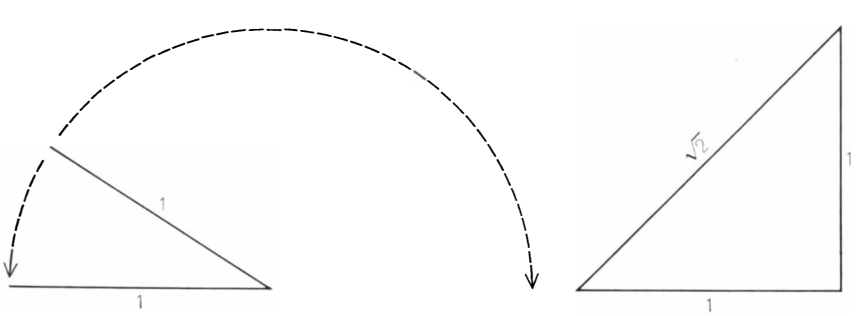
Go-moku problem: O *to play and win*



Answer to the card problem

second player eventually to form a row of three.

Draw games *are* possible on the order-4 cube, but whether the first player can force a win is not, as far as I know, positively established. (There cannot, of course, be a second-player win because of Nash's proof.) As with go-moku, the first player has a strong advantage and a winning strategy is believed to exist. Many computer programs for this game have been written, but the complexity of play is so enormous that I do not think a first-player win has yet been rigorously demonstrated. About a dozen readers have sent me what they consider winning strategies, but detailed formal proofs are still unverified. Most of the strategies involve first taking four of the eight central cells and then proceeding to a forced win. Virtually nothing is known about three-dimensional games where counters are allowed to move from cell to cell.



Answer to the triangle problem

Another unexplored type of 3-space game is one in which two players alternately draw from a limited supply of unit cubes of two or more colors to build a larger cube with some winning objective in view, for example, using cubes of n colors and trying to get a row, on an order-n cube, in which all n colors appear. For such games gravity imposes restraints, since cubes cannot be suspended in midair.

Because drawn games of standard ticktacktoe are possible in 2-space on an order-3 board, and in 3-space on an order-4 board, it was once conjectured that in a space of n dimensions the smallest board allowing a draw was one with $n + 1$ cells on a side. It turned out, however, that although in n-space a board of order $n + 1$ or higher always allows a draw, it is sometimes possible for an n-space board of fewer than $n + 1$ cells on a side to allow a draw. This was first established about 1960 by Alfred W. Hales, a mathematician now at the University of California at Los Angeles, when he constructed a draw on the order-4 hypercube, or fourth-dimension cube. (I am indebted to Golomb for this information.)

Several readers have sent informal but probably valid proofs that the first player can always win on the order-4 hypercube. Whether or not he can force a win on the order-5 hypercube is yet another of the many unanswered questions about extensions and variants of what most people, like Tom Sawyer, regard as a "simple" game.

The answers to last month's rather deceptive quickie problems follow:

1. The smallest number of soldering points remains eight no matter how wires are bent. Because an odd number of edges meet at each corner of a cube, every point requires soldering.

2. Do not put Descartes before the horse.

3. You can evade check forever. Head toward the board's center, always moving your king to a color opposite to that of the knight. Since a knight changes the color of its cell at every move, whenever the king is on a color different from the knight's, no knight move can check the king. Your only danger lies in being trapped in a corner where you can be forced to move diagonally and be checked by the knight's next move.

4. The highest constant is 36 [*see top illustration at left*].

5. One answer: "The value of n is less than one million."

6. He makes twice as much money.

7. The black curve divides the plane

into a number of regions. Trace a round trip along the red curve and it is obvious that every region you enter you must also leave or you will never get back to where you started. Since each entrance and exit is a pair of crossing points, the total number of such spots must be even.

8. 2.3.

9. Two and a half times as high.

10. With one unit side as base and the other unit side free to rotate [*see bottom illustration on opposite page*], the triangle's area is greatest when the altitude is maximum. The third side will then be the square root of 2.

11. $1 + 2 + 3 = 1 \times 2 \times 3$.

12. Only two of the eight possible combinations of crossings create a knot, making the probability of a knot $2/8 = 1/4$.

13. House.

14. The first number, 1,324, raised to any power must end in 6 or 4. The other two numbers, 731 and 1,961, raised to any power must end in 1. Since no number ending in 6 or 4, added to a number ending in 1, can produce a number ending in 1, the equation has no solution.

15. Underground.

16. One proof that the probability is 1/2: Suppose the man has a *Doppelgänger* directly opposite him on the other side of the Pentagon's center and the same distance away. If either man sees three sides, his double must see only two. Since there is an equal probability that either man is at either spot, the probability is 1/2 that he will see three sides.

17. HOBO.

18. The four possible true-false combinations for the two statements are *TT*, *TF*, *FT* and *FF*. The first is eliminated because we were told that one statement is false. The second and third are eliminated because in each case, if one person lied, the other cannot have spoken truly. Therefore both lied. The boy has red hair, the girl black hair.

19. The top illustration at the right shows the two solutions.

20. If $ABCD = 1{,}234$, it is impossible to obtain a sum as large as 12,300. If $ABCD = 3{,}456$, it is impossible to obtain a sum as small as 12,300. Therefore $ABCD = 2{,}345$, from which it is easy to determine that the four dots stand for 4,523.

21. It is well known that every prime greater than 3 is one more or one less than a multiple of 6. It is easy to see that every number of the form $6n \pm 1$ must fall on the same diagonal, therefore the diagonal is certain to catch every prime.

22. Any two positive integers.

23. There are three permissible combinations of true and false for the three statements: *TFF*, *FTF* and *FFT*. The only noncontradictory combination is *FTF*, which means that Feemster owns no books at all.

24. Each month has 11 ambiguous dates (a date such as 8/8/71 is not ambiguous), making 132 in all.

25. A square manhole cover, turned on edge, could slip through its hole and fall into the sewer.

26. Ten digits can be permuted in $10! = 3{,}628{,}800$ different ways. A 10-digit number cannot start with zero, so that we must subtract $3{,}628{,}800/10 = 362{,}880$ to obtain the answer: 3,265,920.

27. No, because after 72 hours it would have been midnight again.

28. "To be or not to be, that is the question."

29. Instead of inscribing the hexagon as shown last month, turn it to the position shown in the middle illustration at the right. The gray lines divide the larger hexagon into 24 congruent triangles, 18 of which form the smaller hexagon. The ratio of areas is $18 : 24 = 3 : 4$, and so if the smaller hexagon has an area of three, the larger one has an area of four.

30. Smith was born in 1892. He was 44 in $44^2 = 1936$.

31. I write "10." This is *any* base written in that base system's notation.

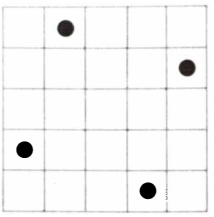
32. U Thant.

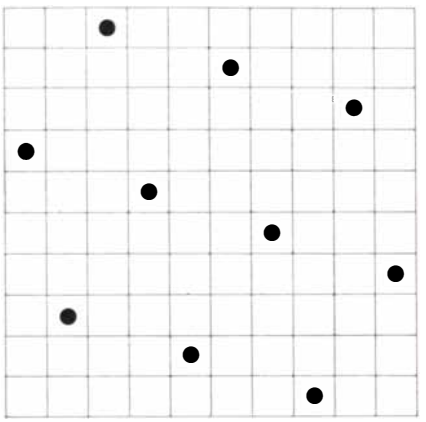
33. Twenty white cubes can abut the red cube. Arrange seven white cubes as shown in the bottom illustration at the right. The red cube goes on top as indicated. Seven more white cubes, in the same pattern and position as the first layer, form layer No. 3. In between layer No. 1 and layer No. 3 six more white cubes can be placed: two on each of two opposite sides of the red cube and single cubes on the remaining two sides.

34. The consecutive letters *RSTU* will spell "rust" or "ruts."

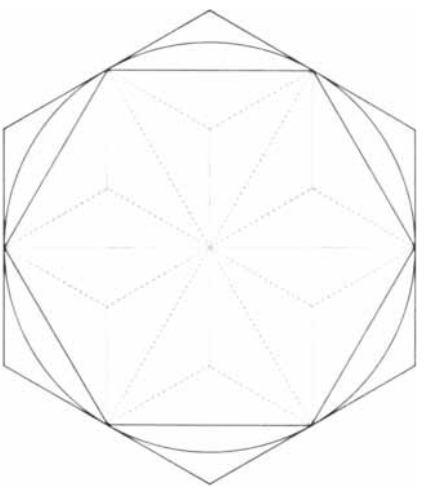
35. He ties one end of the rope to the tree at the edge of the lake, walks around the lake holding the other end of the rope and ties that end to the same tree. The doubled rope is now firmly stretched between the two trees, making it easy for the man to pull himself through the water, by means of the rope, to the island.

36. The dog can be at any point between the boy and the girl, facing either way. Proof: At the end of one hour, place the dog anywhere between the boy and the girl, facing in either direction. Time-reverse all motions and the three will return at the same instant to the starting point.

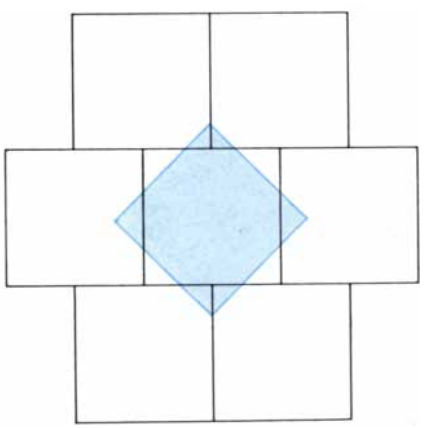




Superqueen solutions



Calculating hexagon areas



Arrangement of the cubes

THE AMATEUR SCIENTIST

How to build a machine to produce low-energy protons and deuterons

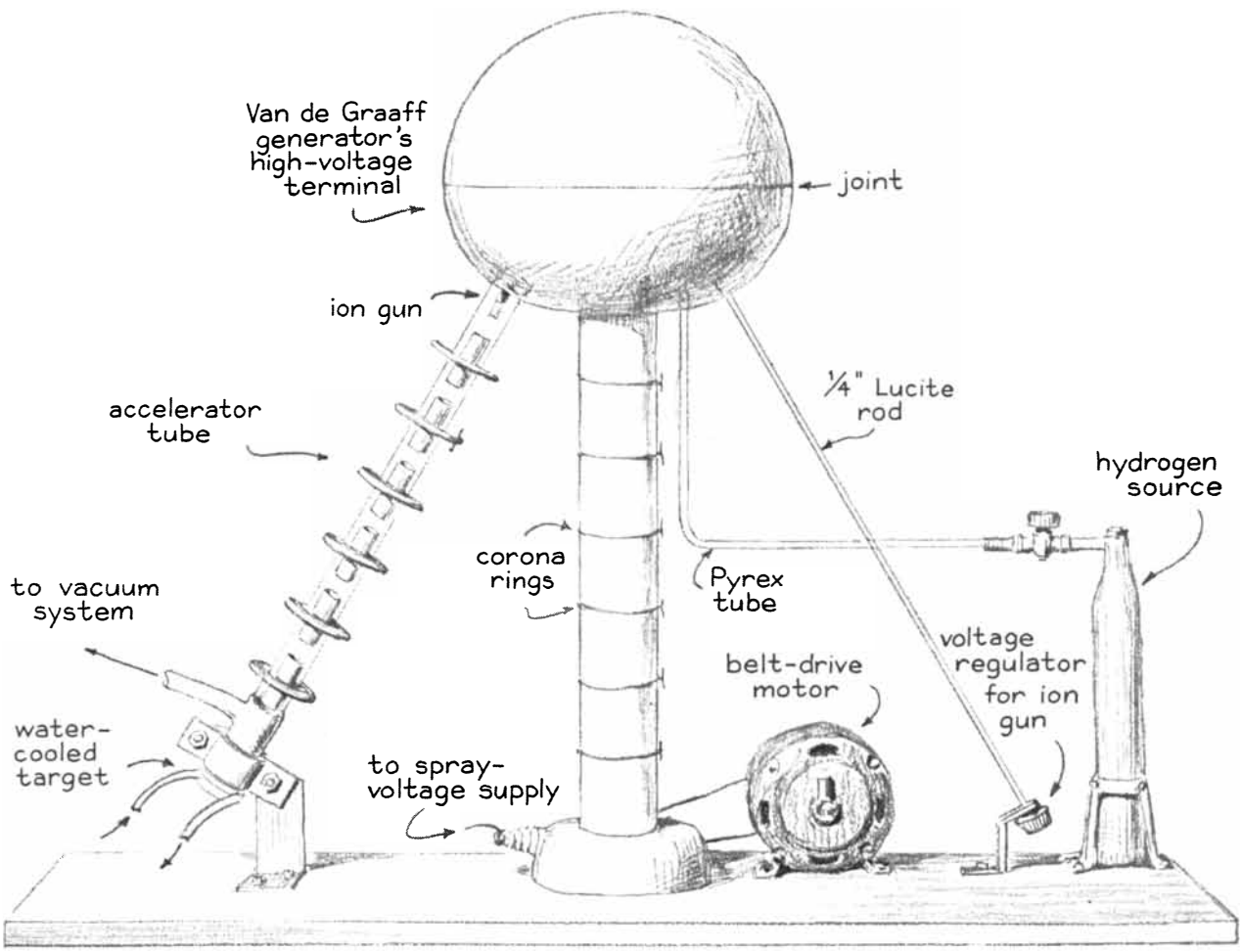
Conducted by C. L. Stong

Although radioactive materials for experimental use can be bought in small amounts without a license, some amateurs have made their own with an apparatus that generates a beam of protons. The protons react with the atomic nuclei in a target to liberate neutrons. The neutrons in turn react with the nuclei in another substance, giving rise to a radioactive isotope.

Small machines that can be made at home do not produce much radioactive substance, but a little goes a long way. An amateur who has made radioactive substances of several kinds is Larry Cress of Pennington Gap, Va. He explains how to build a small proton accelerator and describes experiments that can be conducted with it.

"My accelerator," Cress writes, "consists essentially of a series of electrically insulated copper tubes supported in axial alignment inside a cylindrical vacuum chamber. Hydrogen gas, admitted at one end of the structure, is ionized by a pair of electrodes. The resulting protons (hydrogen nuclei) are accelerated to a target at the other end of the structure by an electrostatic field of high potential. Portions of the field that lie between adjacent ends of the copper tubes act as electrostatic lenses that focus the protons into a beam.

"The accelerating potential is developed by a small electrostatic generator of the Van de Graaff type [*see illustration below*]. Hydrogen is provided by a tank of compressed gas. Gas pressure in the accelerator tube is controlled by a pair of oil diffusion pumps that operate



The proton accelerator built by Larry Cress

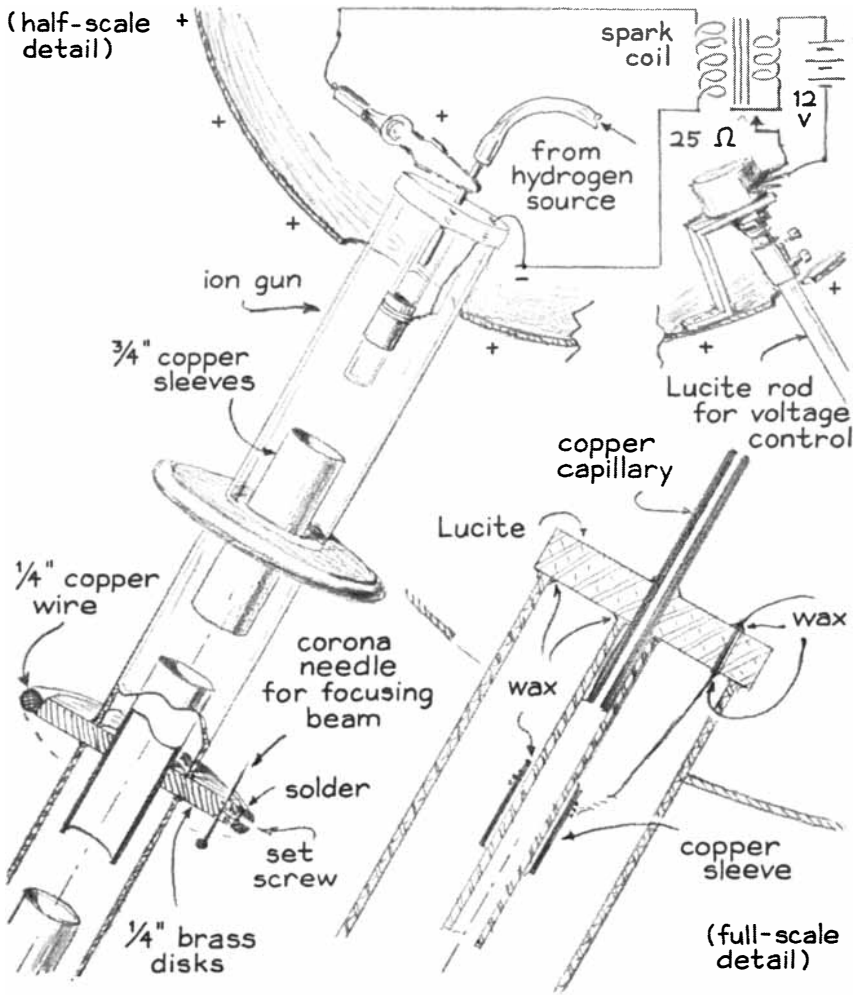
in series and exhaust into a mechanical air pump.

"The accelerator is quite versatile. For example, by substituting deuterium (the heavier isotope of hydrogen) for protons one can accelerate deuterons (deuterium nuclei, consisting of one proton and one neutron). Indeed, most of my experiments have been done with deuterium. By closing the target end of the tube with a window of aluminum foil at ground potential and omitting the hydrogen or deuterium, one can use the apparatus to accelerate a beam of electrons that is accessible in the air beyond the window. The electron beam can promote a host of chemical reactions.

"My Van de Graaff generator was bought as a kit from Morris and Lee (1685 Elmwood Avenue, Buffalo, N.Y. 14207). The generator is rated at a potential of a half-million volts at 20 microamperes. The load imposed by the accelerator tube reduces the effective potential to about 250 kilovolts. The belt that conveys charge to the high-potential electrode of the machine is driven by a 1/4-horsepower induction motor.

"To increase the available current and control the polarity and potential of the output voltage of the generator I spray charge on the belt, at a point close to the bottom of the machine, as a corona from a charged needle point. Potential for energizing the needle is developed by an induction coil of the Ford Model-T type that operates from a six-volt transformer. The output of the induction coil is converted to direct current by a vacuum diode of the type used in small X-ray machines. The potential of the spray voltage is controlled by a 25-ohm rheostat in series with the primary winding of the induction coil. The charging current is measured with a microammeter in series with the grounded side of the coil. The rectifier is not strictly necessary because the output of induction coils tends to be unidirectional, but the inclusion of the diode decreased the fluctuation of the output voltage and increased the current.

"Electrical connection between the accelerator tube and the high-voltage terminal of the Van de Graaff machine can be made either by fitting the end of the accelerator tube with a duplicate of the high-voltage terminal and placing the two terminals in contact or by inserting the end of the accelerator tube into the high-voltage terminal of the generator. I elected to use the latter scheme, primarily because large hollow terminals of spun aluminum are costly. The high-voltage terminal consists of an unperforated hemisphere that mates



Details of the accelerator tube

with a second hemisphere curving smoothly inward to an opening that admits the charged belt.

"I opened the terminal at the bonding joint of the hemisphere, so that the apparatus for ionizing the hydrogen gas could be placed inside. A hole 1½ inches in diameter was made in the lower surface of the perforated hemisphere to admit the accelerator tube. The edge of the hole was rounded inward and smoothed to prevent sparking or corona discharge, which tends to occur at sharp projections.

"Both the terminal and the belt are supported by a cylindrical insulating column. Rings consisting of a single turn of bare copper wire were placed around the column at intervals of three inches. Known as corona rings, the wires distribute voltage uniformly along the column and prevent direct discharge down the column to ground.

"The assembly of the machine and the modification of the high-voltage terminal completes the easy part of the job. The hard part consists in making a gas-tight accelerator tube by sandwiching electrode assemblies between lengths of 1¼-inch Pyrex tubing and closing the ends. The tube must be capable of retaining a pressure of 10^{-5} torr.

"The glass is cut into three-inch lengths by filing a nick completely around the tubing and touching the nicked portion to a Nichrome wire that is electrically heated to redness. The glass will crack at the point of contact. By slowly rotating the tubing while it is in contact with the hot wire, lead the crack completely around the glass. If the glass does not part immediately, exert a strong pull on the ends.

"Each electrode assembly includes a circular brass plate 2½ inches in diameter and 1/4 inch thick. Drill a 3/4-inch hole accurately through the center of each disk and through it insert a 2¼-inch length of straight copper tubing of 3/4-inch outside diameter. The edges of the cut ends of the tubes must be rounded and preferably polished to minimize corona discharge. The edge of each disk must be similarly rounded. If the experi-

menter does not have access to an engine lathe, the edges can be rounded by soldering a ring of 1/4-inch copper wire to the perimeter of each disk.

"The copper tubes must be supported in axial alignment by the glass sections. To make the assembly I clamped a dowel vertically in a vise. A *T* fitting of plastic pipe 1¼ inches in diameter was slipped over the dowel. The fitting rested on the vise. One of the copper tubes, with its brass disk, was slipped over the dowel and lowered so that the disk was centered with respect to the *T* fitting and rested on the upper end of the *T*. The joint was sealed with Apiezon W-100 vacuum wax. The melted wax was applied with a medicine dropper after the parts had been warmed with a small flame. A glass tube was then slipped over the dowel, centered on the brass disk and sealed in place. The remaining five electrode elements of the tube were assembled in the same way. The completed structure was lifted from the dowel after the wax hardened.

"The ion source, which closes the high-voltage end of the accelerator tube, consists of another identical length of Pyrex tubing capped at one end by a disk of clear plastic 1/4 inch thick that supports a straight length of copper capillary tubing [*see illustration on preceding page*]. Tubing of the kind used for the pilot light of a gas stove is adequate. An insulator of glass tubing of any convenient diameter surrounds the capillary and extends about 1/2 inch beyond the copper. Around the glass tube is a metal sleeve. The lower edge of the sleeve is positioned about 3/4 inch from the end of the glass tube.

"The capillary admits hydrogen or other gases to the accelerator tube and also serves as the positive electrode of the ionizing potential. The metal sleeve is the negative electrode and is supplied with an uninsulated wire of solid copper that extends through the plastic cap. Before sealing the ion source to the accelerator tube, install an adjustable needle point near the edge of the brass disk that is next to the high-voltage electrode. The needle is used to focus the beam.

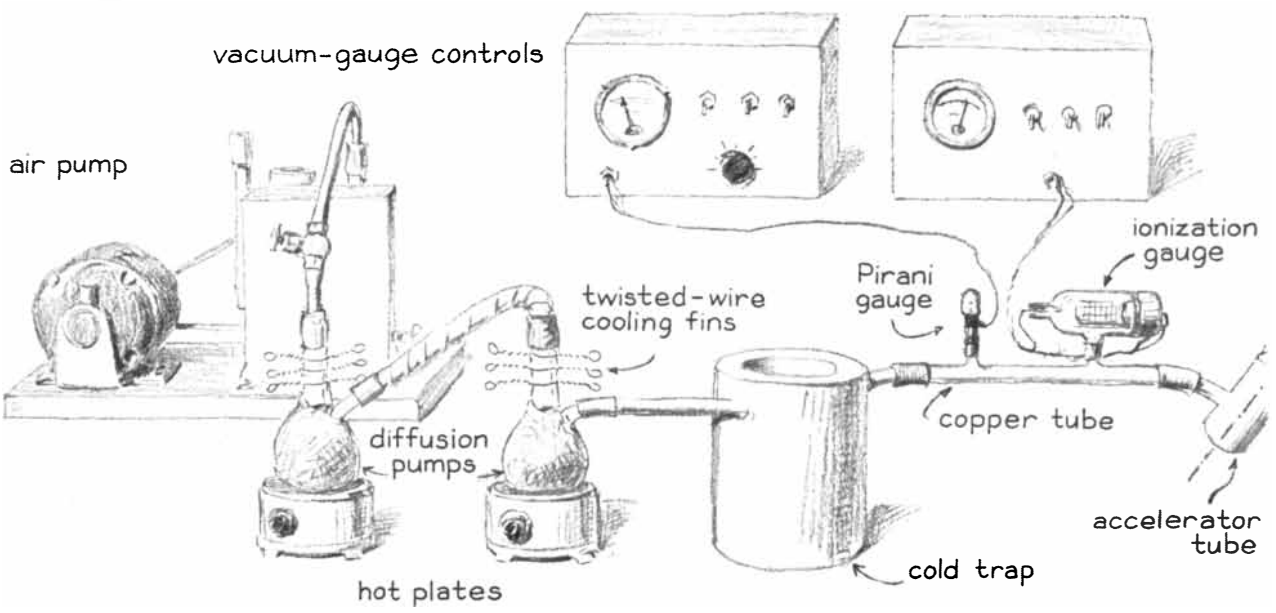
"Energizing potential for the ion source is provided by an induction coil of the Model-T type. Any comparable coil that develops a potential of at least 10 kilovolts can be used. The coil, a pair of six-volt dry batteries and a 25-ohm rheostat are connected in series and installed inside the high-voltage terminal of the Van de Graaff generator. These parts should be spaced at least an inch from the belt and the upper pulley. Access to the rheostat, which also functions as an on-off switch, is gained by removing the upper half of the high-voltage terminal. Alternatively the rheostat can be operated by a plastic rod that passes through a small hole in the high-voltage terminal.

"As mentioned above, the output of induction coils tends to be unidirectional. The negative terminal of the coil is connected to the metal sleeve of the ion source; the positive terminal, to the copper capillary. To determine the polarity of the coil connect its output to a gas-discharge tube that has been exhausted to a pressure of about one torr. At this pressure the glowing gas will separate from one of the electrodes, creating a dark space. This effect, known as the Faraday dark space, identifies the negative electrode.

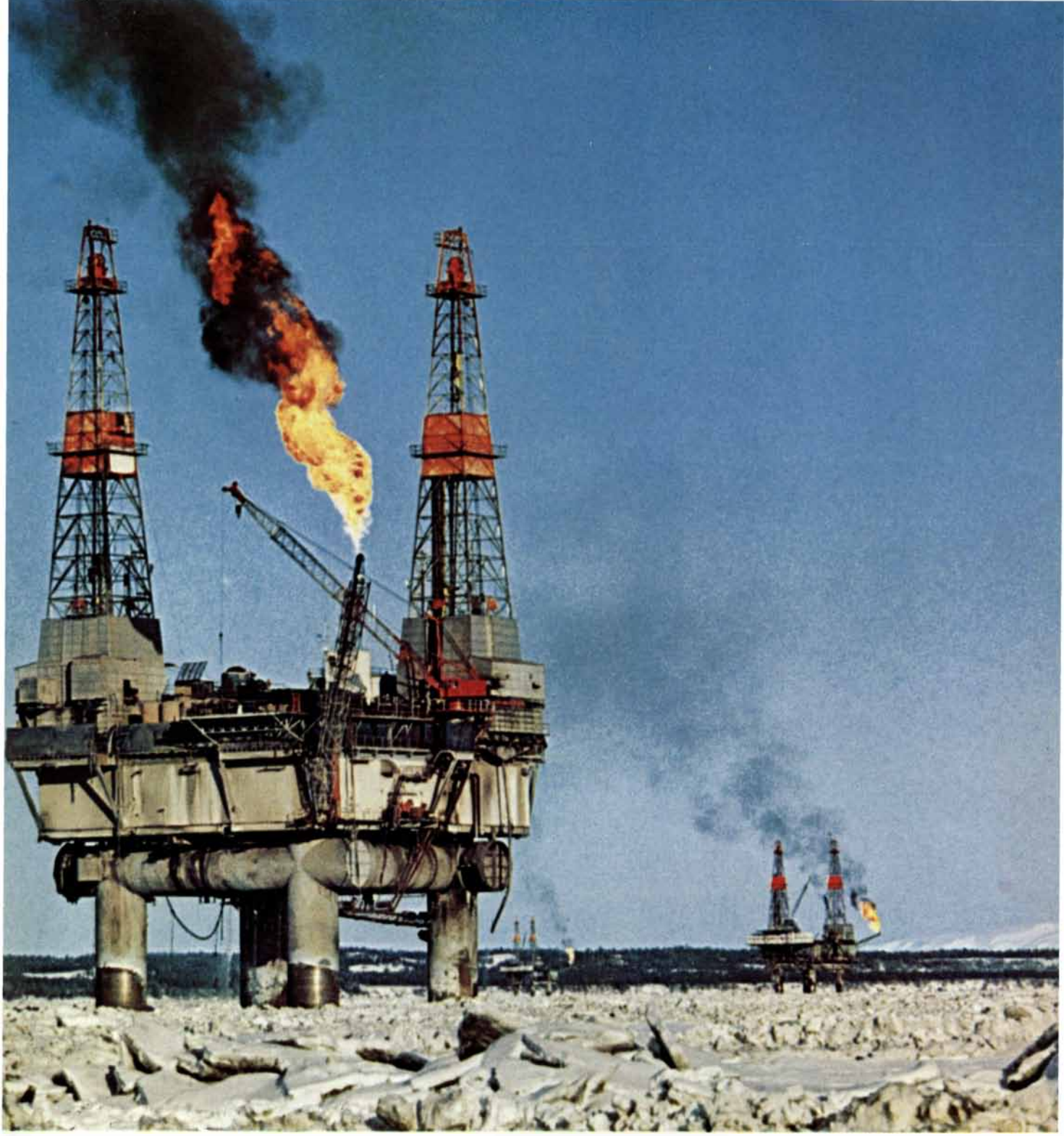
"An appropriate gas-discharge tube can be improvised by plugging the ends of any glass tube about 15 centimeters long with rubber stoppers. Pierce one stopper with a nail and the other with a short length of copper capillary. The nail functions as one electrode; the copper capillary serves both as an electrode and as a port through which air is pumped from the tube. The pressure can be reduced to one torr with a mechanical air pump. The completed accelerator tube is supported at an angle of about 60 degrees by the high-voltage terminal atop the Van de Graaff machine and by an improvised clamp at the bottom, which is anchored to the base of the apparatus.

"The vacuum system consists of two single-stage Hickman oil diffusion pumps and a mechanical air pump connected in series. Before I bought the mechanical pump I used a pair of refrigerator compressors in reverse. The compressors developed the required low pressure but overheated after several hours of continuous use. Some four hours of continuous pumping are needed to exhaust the accelerator tube from atmospheric pressure to a pressure of 10^{-5} torr.

"The diffusion pumps were obtained from Morris and Lee. I use butyl phthalate as the pumping fluid and heat it with small laboratory hot plates oper-



Vacuum system of the accelerator



Off-shore platforms, Cook Inlet, Alaska

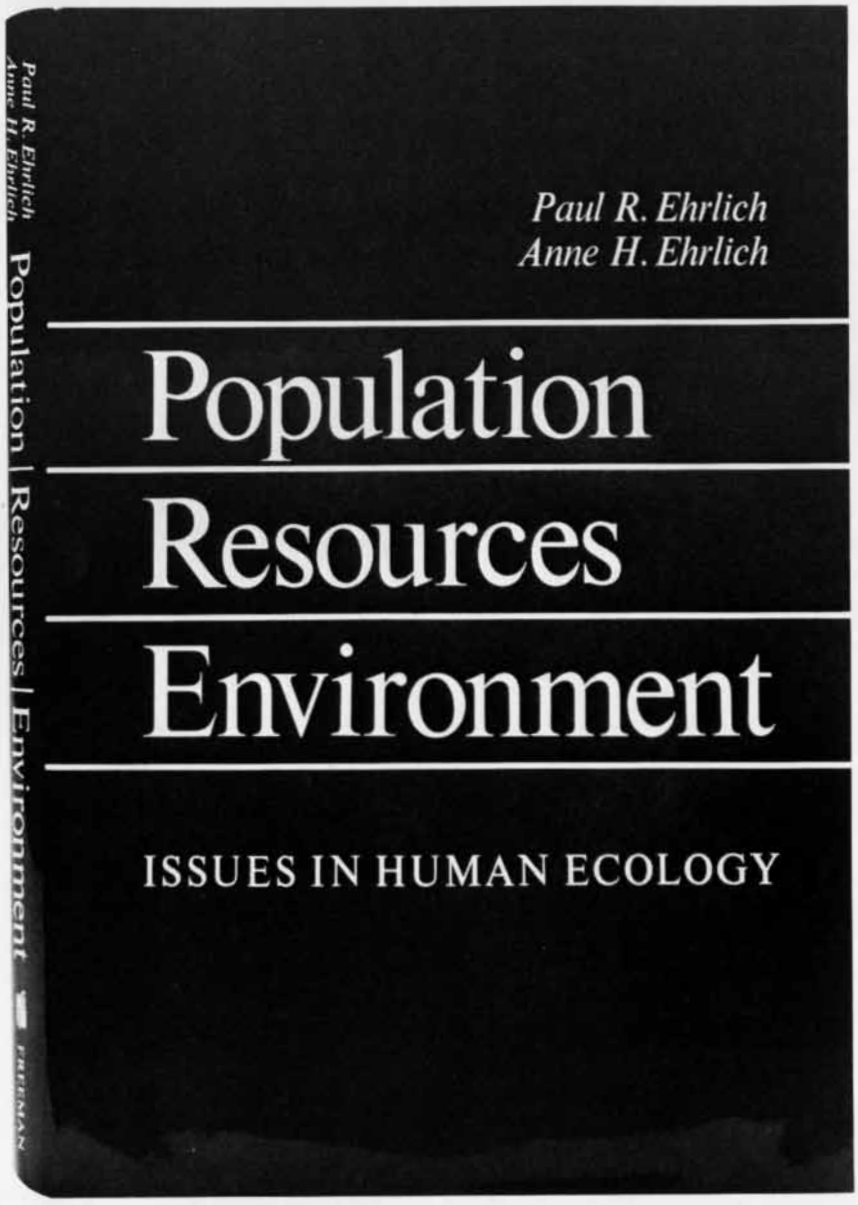
JOSEPH RYCHETNIK/ BLACK STAR

The September issue of **SCIENTIFIC AMERICAN** will be devoted to

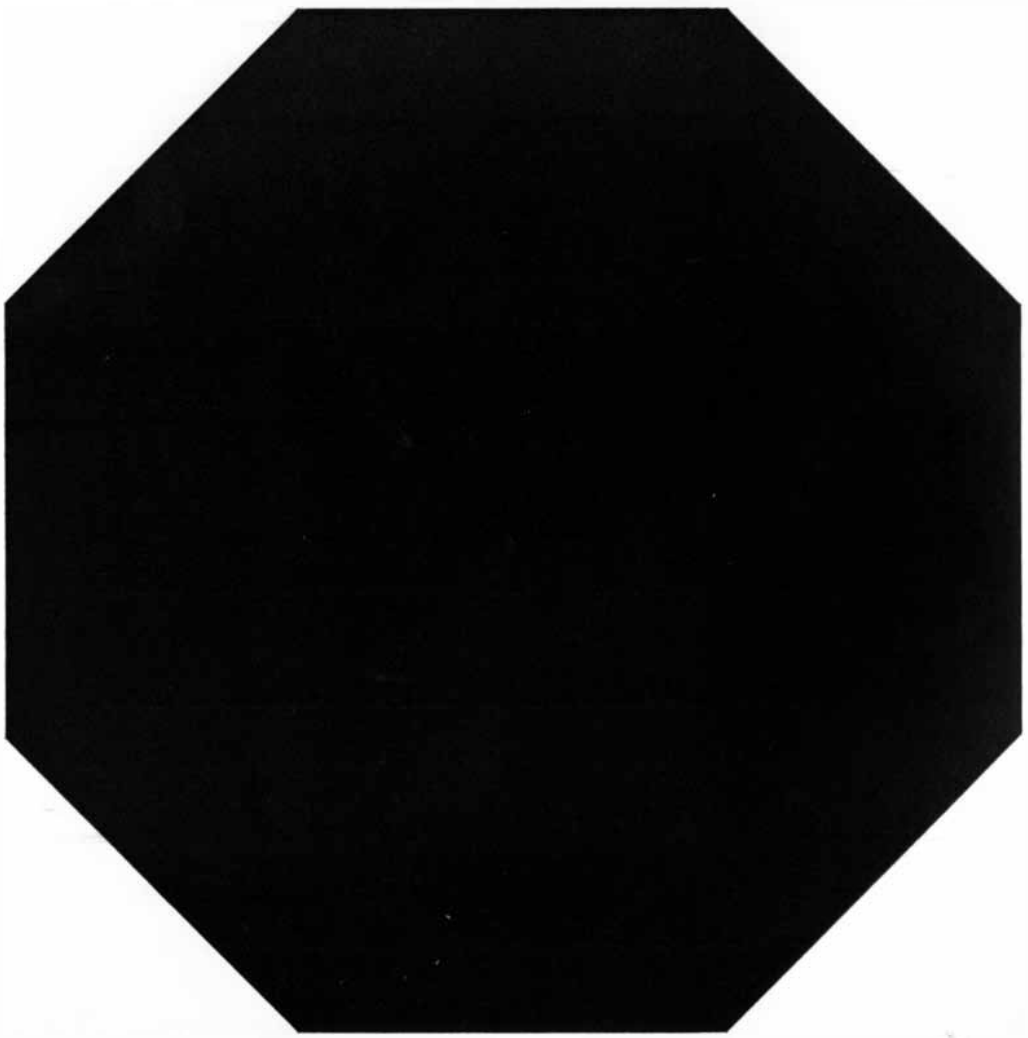
ENERGY and POWER

"The Ehrlichs will prevail because they are essentially right."

—New York Times Book Review



This is how most oil companies look at things.



This is how we look at things.

Most people see a good thing and hang onto it.

We see it, and we have to try and make it better.

Every oil company has industrial lubricants they can be proud of.

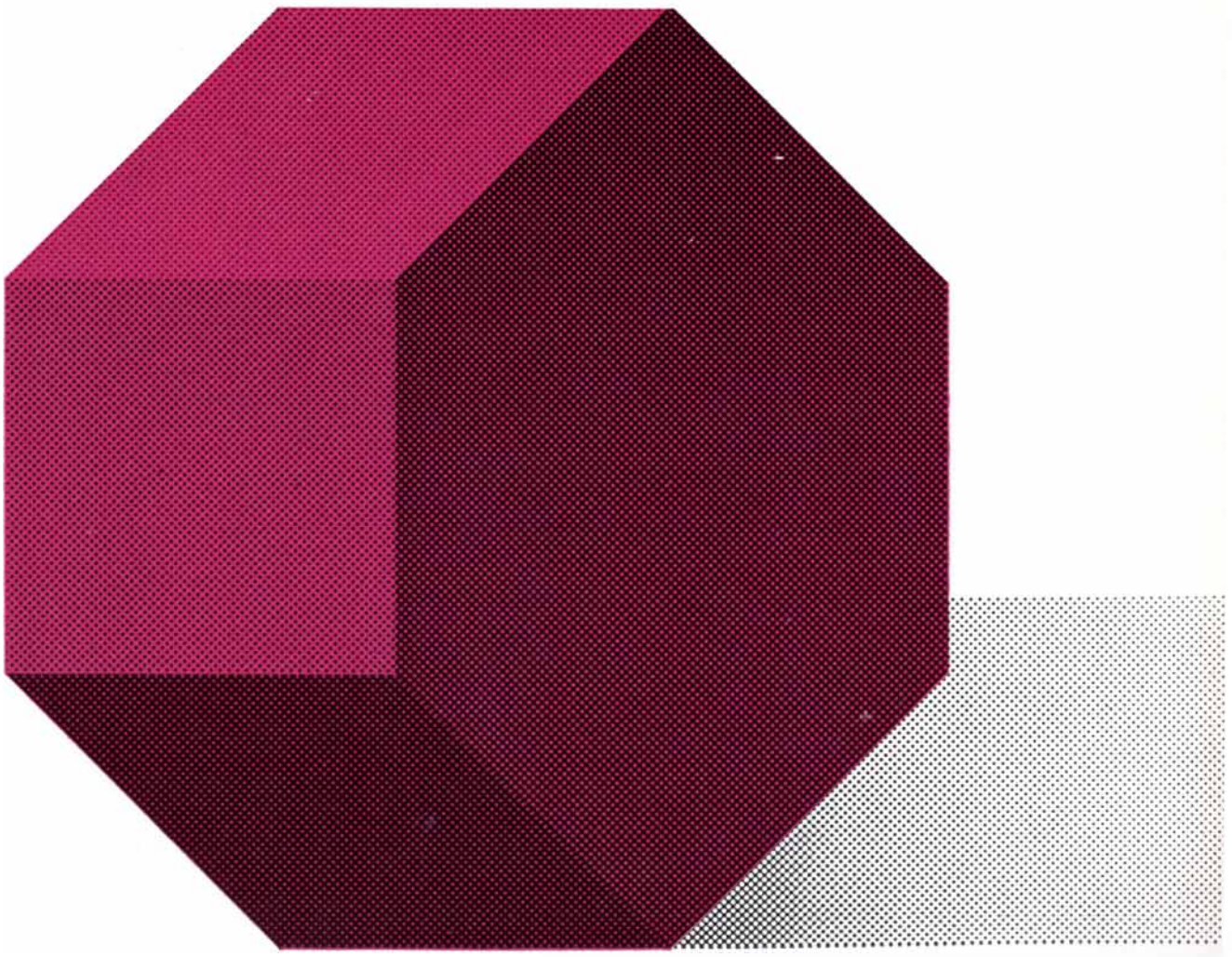
And so do we.

But we have to keep coming up with lubricants for industry that are newer and better.

That's why we're important to you.

AMERICAN

American Oil Company



ated at a temperature of 350 degrees Fahrenheit. A cold trap, consisting of a copper coil immersed in a slurry of dry ice and alcohol, must be installed between the outlet of the accelerator tube and the inlet to the pumps. The cold slurry is contained in a pint-sized tin can insulated with a thick layer of Styrofoam. The pumps cannot reduce the pressure to the required 10^{-5} torr without the cold trap. To prevent ice from forming in the trap I do not add the slurry until the system has been pumped to 10^{-2} torr.

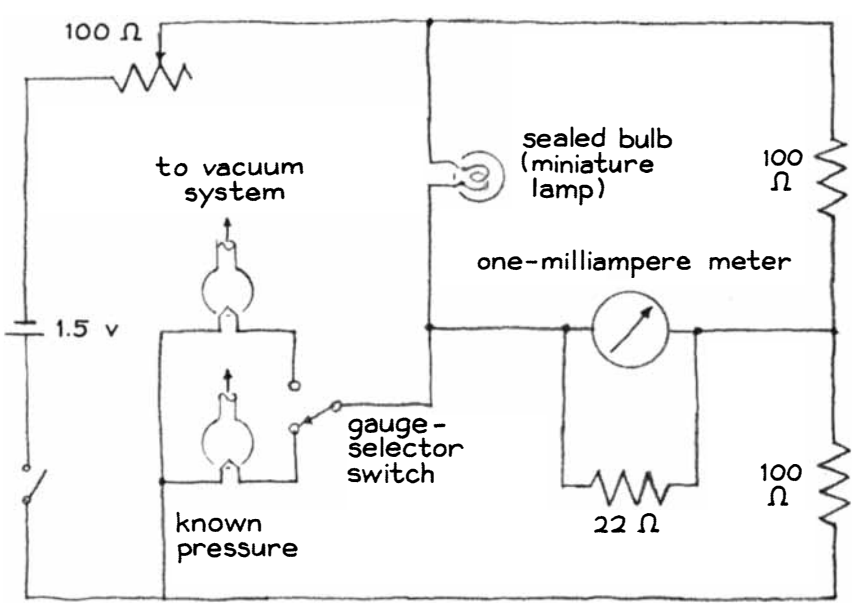
"Pressure in the accelerator tube can be measured down to 10^{-3} torr with a vacuum gauge of the Pirani type. The pressure-sensing elements of the gauge consist of three miniature incandescent lamps of identical resistance, which are rated at six volts. I checked the resistance of several lamps with a Wheatstone bridge and selected the most closely matched set. The resistance of each lamp should be about 15 ohms.

"The glass bulbs of two of the lamps must be opened for connection to the vacuum system. Heat the bases in a small gas flame to unsolder the filament leads and soften the cement that bonds the metal to the glass. Pull the base off to expose the tip of the tube through which the lamps were exhausted. Heat the glass tip to redness in the edge of a gas flame and touch it with a wet cloth. The glass will fracture, admitting air to the bulb slowly. After five minutes pick off the fractured tip with a pair of tweezers. Attach a short length of seven-millimeter glass tubing to the opening with vacuum wax. Wire the three bulbs into the Wheatstone bridge circuit.

"One of the opened bulbs is used to calibrate the gauge. Connect the bulb to one arm of a glass *T*. To the *T* also connect the vacuum system and a McLeod gauge. An appropriate McLeod gauge can be obtained from Morris and Lee, or it can be made [see "The Amateur Scientist"; SCIENTIFIC AMERICAN, December, 1965]. To calibrate the Pirani gauge, switch on the bridge circuit. Operate the two-position switch to connect the calibration bulb into the circuit. Adjust the 100-ohm rheostat for a meter indication of one milliampere. Start the vacuum pumps and, by operating the needle valve, exhaust the system to a series of measured pressures down to 10^{-3} torr. Tabulate each pressure as measured by the McLeod gauge and simultaneously record the corresponding meter reading. Draw a calibration graph by plotting the known pressures against the meter readings.

"One could dispense with the Pirani gauge and measure the pressure of the accelerator tube directly with the McLeod gauge. Although the McLeod gauge indicates absolute pressure and is universally used for calibrating other instruments, it is inconvenient to operate. The effort expended in making the Pirani gauge is amply repaid by the convenience it affords.

"The remaining opened bulb of the Pirani gauge is connected in the vacuum system. The calibration bulb is normally at atmospheric pressure. It is used to set the meter at one milliampere when the other opened bulb is under vacuum. The current should be checked with the calibration bulb before each pressure measurement.



Circuitry of the Pirani gauge

"For measuring pressures below 10^{-3} torr I use an ionization gauge that has a G71–2 triode. This tube is manufactured by the High Vacuum Corporation (Hingham, Mass. 02043). The gauge response is assumed to be linear from 100 microamperes at a pressure of 10^{-3} torr to one microampere at 10^{-5} torr. The mechanical air pump exhausts the system to 10^{-2} torr in about 30 minutes. After this pressure is attained I turn on the heaters of the diffusion pumps. If the system has been opened to the atmosphere, the diffusion pumps require several hours to reduce the pressure to 10^{-5} torr. They do the job in about 30 minutes, however, if the system is maintained at a reasonably good vacuum. I try to keep the system under vacuum at all times. It must of course be opened occasionally to change the target.

"Hydrogen or another gas is admitted to the copper capillary of the ion source from a Pyrex tube that connects through a needle valve and a pressure-reducing valve to the storage tank. The induction coil ionizes some of the gas as it emerges from the capillary. Excess gas is continuously removed from the accelerator tube by the pumps. If the rate at which gas is admitted exceeds the rate at which the pumps can remove the excess, pressure rises in the accelerator tube, as indicated by the ionization gauge. At a sufficiently high pressure both the accelerator tube and the Pyrex tube that leads to the storage tank emit the characteristic glow of a gas-discharge tube. The rate of flow of hydrogen or deuterium must be adjusted accordingly.

"Several difficulties may arise when the apparatus is put in operation. The tubing may leak at the joints, as indicated by the failure of the pumps to exhaust the system to 10^{-5} torr. The detection of leaks is an art in itself [see "The Amateur Scientist"; SCIENTIFIC AMERICAN, February, 1961]. I shall mention only one detection technique. Spray the wax joints one at a time with rubbing alcohol. If there is a leak, the pointer of the ionization gauge usually jumps slightly. Reheat the wax to seal the leak.

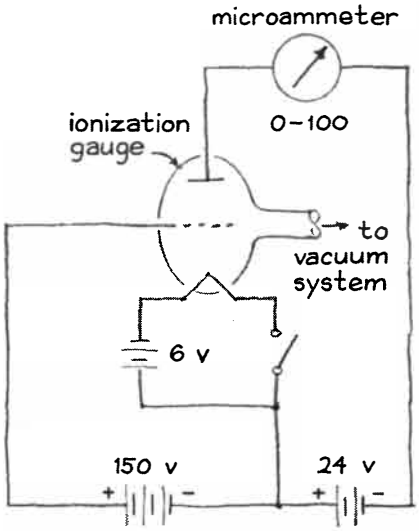
"Gas will be liberated from the internal surfaces of the system during pumpdown, particularly when the beam of accelerated particles strikes the electrodes or other surfaces. When the pressure has fallen to 10^{-5} torr, start the Van de Graaff generator and gradually increase the output voltage by regulating the amount of charge sprayed onto the belt. Watch the ionization gauge. Do not increase the voltage at a rate that causes the pressure to rise above 10^{-4} torr.

"When the pressure remains constant at 10^{-5} torr with an applied potential of 250 kilovolts, focus the proton beam. The beam can be detected by inserting a thin plate of quartz over the target. The quartz fluoresces, emitting a spot of purple light in the area struck by the beam. Advance or retard the adjustable needle in the brass disk of the accelerator tube nearest the terminal to focus the beam. The spot should focus to a diameter of about one millimeter. The quartz plate is mounted on an improvised pivot so that it can conveniently be swung away from the target when it is not in use.

"The accelerating potential of 250 kilovolts develops a proton beam of relatively low energy compared with the beams of full-scale accelerators. For this reason the amateur experimenter is limited to the investigation of nuclear reactions in elements of low atomic weight. The bombardment of boron and fluoride targets with protons liberates intense gamma rays with energies substantially above 250 kilovolts. The rays can be detected with a Geiger-Müller counter or a sodium-iodide scintillator having a photomultiplier tube and a meter circuit. As the potential of the Van de Graaff generator is gradually increased from zero, the emission of gamma rays from boron is detected at 163 kilovolts and from fluoride at 224 kilovolts.

"Most of my experiments have been conducted by bombarding deuterons with deuterons. The reaction liberates neutrons. Neutrons combine with many elements, which become radioactive as a consequence of the reaction. The elements can be identified by the nature of their subsequent emissions. Neutron absorption also causes some elements, notably uranium, to fission. The half-life of the resulting radioactive products ranges from minutes to millenniums. An appropriate target of uranium can be bought from A. D. Mackay, Inc. (198 Broadway, New York, N.Y. 10038) for $5. The order must be accompanied by a self-addressed envelope.

"The energy levels of radioactive nuclei can be determined by measuring the energy of the emitted gamma rays. A good target for this reaction is a thin film of heavy water (D_2O). I freeze a thin film of heavy water on the inner surface of a copper target by cooling the exterior surface with liquid nitrogen. Liquid nitrogen can be obtained from dealers in welding supplies. Another satisfactory target is deuterium orthophosphate, which can be made by adding heavy water a drop at a time to a small amount of phosphorus pentoxide until



Circuitry of the ionization gauge

all the material has reacted. The energy of neutrons liberated by the deuteron-deuteron reaction averages about 2.6 million electron volts. At a Van de Graaff potential of 200 kilovolts and a current of 10 microamperes the reaction will liberate approximately 5×10^7 neutrons per second. I prepared for the experiments by buying a 25-liter cylinder of deuterium and 100 grams of heavy water from Bio-Rad Laboratories (22 Jones Street, New York, N.Y. 10014).

"A word of warning is appropriate. Although the accelerator is small and the potential of the Van de Graaff machine is relatively harmless, the proton beam and the products of the nuclear reactions are hazardous. In addition to emitting gamma rays, the machine can generate X rays of substantial intensity. The apparatus must be effectively shielded. For shielding I use a double layer of solid concrete blocks 18 inches thick. In addition I surround the target with several blocks of paraffin when experimenting with targets that emit neutrons. I monitor the control side of the shielding at all times with a dosimeter and wear a film badge when working with the accelerator or its products. Dosimeters and film badges can be obtained from the R. S. Landauer Company (Glenwood, Ill. 60425). The experimenter should bear in mind that the U.S. Atomic Energy Commission requires that personnel working with radiation materials receive not more than 100 millirems of radiation per 40-hour week.

"I shall be glad to answer questions from those who undertake the construction of the accelerator. Correspondence can be addressed to Larry Cress, R.F.D. No. 1, Pennington Gap, Va. 24277."

BOOKS

A compendium of curious events reported by the Center for Short-Lived Phenomena

by Philip Morrison

ANNUAL REPORT 1970, by the Center for Short-Lived Phenomena. Smithsonian Institution, Cambridge, Mass. ($5). On April 8 an early-rising resident of West Hartford, Conn., saw "a bright streak like a bomb bursting." Later, at 6:00 A.M., Paul J. Cassarino of suburban Wethersfield awoke to find plaster on the floor of his living room. A meteorite weighing three-quarters of a pound had fallen through the plywood roof to lodge halfway through the ceiling!

Something of Cassarino's sense of celestial intimacy was transferred by postcard within a day or two to the many thousands of friends and subscribers of the service of the Center for Short-Lived Phenomena. The card told the facts, as telephoned to Cambridge by the very men who identified the rock as a *bona fide* meteorite. (It was a common kind of chondrite, and it was in the hands of expert analysts before the week was out.)

The Center for Short-Lived Phenomena has been at work since January, 1968. It grew out of efforts to "catch a falling star"—more prosaically to bring a freshly fallen meteorite into the laboratory quickly enough so that information about the cosmic ray intensity in space near the earth, perishably locked in short-lived radioactivity, would not have fully decayed. It was clear from the start, however, that many kinds of "fast and accurate information on the occurrence of short-lived natural phenomena" would be of interest to scientists in several disciplines. Today the center has 3,000 registered correspondents in some 150 countries; these scientists or organizations report nearby short-lived events and follow up on their first alerts, and in turn they receive the reports of similar events around the world. The chief resource of the center is the global communications network of the Smithsonian Astrophysical Observatory, which uses commercial, NASA and other Government circuits. The center stays in touch with event areas as the events are in progress; it immediately alerts those persons and agencies that might quickly respond, and within 24 hours dispatches postcards that broadcast the alert; then it forwards authentic data and conclusions as study of the event proceeds.

This 300-page book, with maps and photographs, its text in unjustified typescript, indexes the work of the center from its very start. The bulk of the book is a summary of the 113 events in 51 countries that the center reported during 1970. These events fall into four classes. Earth-science events are just under 50 in number; they include 19 major earthquakes, 22 volcanic eruptions, tsunamis, landslides, floods, storm surges and more. Biological events are even more varied; they include oil spills and other major pollution events as imminent threats to the web of life, and they extend to events remote from the hand of man, such as the migrations of animals, blights and epizootics. The astrophysical events are mostly fireballs and meteorite falls (plus recurrent moonquakes!). Archaeology and anthropology claim such occurrences as the discovery of a ruin soon to be flooded or the first contact with a new group of people. The unifying theme is the need for on-the-spot scientific study of the event before its course is run; naturally this interest overlaps with, but is not the same as, the interests of the government agencies all over the world that are generally responsible for resource control, rescue and relief.

Tragedy is present, since everywhere the lives of men remain at the mercy of the largest forces of nature. The terrible November storm surges of the Bay of Bengal were Event No. 100 of 1970; the event was of course intensively reported in the regular news media, but the summary here (by Dr. M. Q. Khuda of Dacca) is eloquent in its simplicity. Although many hundreds of thousands died, "no permanent change in the land mark is visible." Mankind is perhaps most vulnerable along the many mouths of the Ganges. The avalanche that entirely buried Yungay in Peru under some 10 feet of mud in the earthquake of May, 1970, is described with equal care.

Late this summer the inhabitants of the Mississippi Valley from its Canadian head to the Gulf can expect a cyclical peak in the number of migrating monarch butterflies; 1970 was the first of the two peak years that recur on a six- or eight-year cycle. Thousands of the bright-yellow-and-black creatures can be seen on a single tree. The cycle is the result of an interaction of the insect and a virus.

Why do whales beach themselves? Here is the account of the stranding of about 150 false killer whales, animals weighing less than a ton each, on a shelving Florida beach. Towed to sea by boats, they immediately swam back to beach themselves all over again. After a second towing effort perhaps a third of the animals found their way out to sea. On autopsy there was no sign of bacterial disease, but it was noted that the water was abnormally cold.

The crown-of-thorns starfish, that hungry predator on the living polyp of the hard madrepore coral, continues to spread alarmingly over the Pacific, again for uncertain reasons. The notorious Malaysian frog war of last November is reinterpreted; afterward there were tadpoles aplenty. Those 10,000 frogs were making love, not war. Deception Island, a drowned volcanic crater 500 miles south of Cape Horn, blew up in the wintry August of 1970. Its dust and ashes were "probably the most important [for] the last three years or so. It seems that the Dust Veil Index will be somewhere in the range between 150 and 500 (Krakatau 1883 eruption equals 1,000...)." No one was there; the first news came from Antarctic bases some 50 miles away, where a rain of black ash covered the entire area to a depth of millimeters.

By December, 1970, people had visited Deception; a photograph was published recently. The center's main purpose was thus fulfilled. Indeed, more than 84 of the events it reported in 1970 were investigated. But he who stays at

home—amateur, student, teacher—can share a little of the investigator's excitement and awe at the changing fabric of our world by reading the center's alerts. There are services matched to all budgets and all levels of specialization. You can arrange for notice by wire, by airmail card or by weekly or even monthly batches, in each class of events. Schools and museums should be particularly interested. The cards supply just those details of place and time, with cogent appraisals, that the news services often omit.

The performance of this pioneer office is not faultless: the *Annual Report*—with much too bland a title—could be improved by greater editorial resources. It is nonetheless splendid reading as it stands; the "real time" services are even better and the future holds more. Surely the center is the first form of what someday will be a knowing and prescient worldwide public monitoring service both for action and for thought.

Long live the Center for Short-Lived Phenomena!

INTERNATIONAL TELECOMMUNICATIONS AND INTERNATIONAL LAW: THE REGULATION OF THE RADIO SPECTRUM, by David M. Leive. Oceana Publications, Inc. Dobbs Ferry, N.Y. ($16.50). Eye, ear and radio receiver discriminate signals not only in time and space but also in quality: color, pitch or frequency. This new dimension (whose utility depends on the fact that the time scale for meaningful signal changes is much slower than the inner time variation that establishes quality) is the main basis of an important body of international behavior and law. This detailed study by an attorney on international matters for the Communications Satellite Corporation is a critical analysis of the body of law and of the agency that administers it. His study is set in the context of history; it is much concerned with the rise of space communications, a wholly new domain that was the topic of a special international conference this past month.

Frequency bands are allocated by international treaty within three world regions according to types of service. The Plenipotentiary Conference of the International Telecommunication Union holds the supreme legislative power. The ITU is ruled by that conference, which meets every five or six years to work out anew the entire complex treaty, ultimately signed by all the member countries. The main interest of this volume is the work of a board of five independent men elected at the conferences, each man from a different country. These men form the International Frequency Registration Board, an organ with many faces. It issues findings as a court does, and it publicly notes and formally records specific national frequency assignments within the bands as a registry office records deeds to land. It conciliates conflicts, promulgates technical standards and procedures and offers aid to developing countries. It is busy; its 100-member staff, working among several other distinct agencies of the ITU in Geneva, publishes about 50 pages of frequency assignments weekly. The full frequency list, published every year or two as computer printout, gave at the end of 1968 the particulars of about 364,000 individual assignments. These assignments are made by nations in presumed compliance with the treaty and conveyed to the Frequency Registration Board in the hope that acceptance in the Master Register will confer some measure of national claim to the frequency so assigned.

Just what measure of protection actually accrues in various circumstances is the main content of this book, a subtle study of an organization whose practices are more complex than those of the most hairsplitting faculty senate. Nations can make any specific assignments they hold to be proper, that is, to cause no harmful interference. It is up to the board to concur or not, under a shifting and ambiguous regime of precedent and criterion. The board has no powers other than publicity and persuasion. Its sanction is mainly that if its finding is ignored, its protection can less easily be sought by the same party in a future case. True, the nations are bound to obey the frequency allocations, but they need not even do that if no harmful interference results. What is harmful requires judgment; it is plain that priority claims on a frequency can reflect obsolete practice or vary in their effect with the sunspot cycle. The radio spectrum is not consumed; rather, a channel is being wasted whenever it is not in use. Yet the overall arrangement is remarkably self-enforcing; it "offers an excellent example of how national self-interest can be harnessed to promote compliance with international law." Generally administrations would rather have interference-free communication than perfect their legal claims.

There are sharp limits to this rule of reason. All nations have veto power over the Frequency Registration Board. Nations "retain their entire freedom with regard to military radio installations," and such installations are not really under the board (although there is some fine print). Shortwave broadcasting is specially treated; European history has made its regulation intractable. Demands for broadcasting channels far exceed the space; the board publishes only a list of announced schedules for those channels and then washes its hands of the broadcasters—4,000 clamorous occupants crowding into a few hundred voice channels. These days the technologically advanced nations are more and more moving their heavy message traffic by way of cables and satellite links. Point-to-point high-frequency service by way of the ionosphere becomes increasingly the choice of the developing nations. They know that a few thin antenna masts and a houseful of not very costly gear can put them in contact with the world, provided that the frequencies are available. National pressure is therefore greatest within the high-frequency bands, from a few megacycles to 30 megacycles. Hence many smaller nations (to name three that spoke up: Cuba, Syria and Burma) defended the board in 1965 against a European and U.S. proposal to eliminate it in favor of a single director with purely technical duties. The board was cut in size from 11 to five, but it remains an elective and independent body.

Leive devotes a chapter to his recommendations. He comes down on the side of the board and its usefulness, even though he can see its shortcomings. There are many tasks that no mere technician can perform. He wants for the board more powers, more law, less ambiguous procedures and more independence, financial and personal. Perhaps its gentle control could even be extended to assigning exclusive sectors of the valuable orbit above the Equator.

The ITU regime began with a Paris conference in 1865, attended by 20 European states, that adopted the first International Telegraph Convention, some of whose rules survive today. Administration under that agreement was vested until 1948 in the Berne Bureau, composed of Swiss personnel and financed by the Swiss government! Those were indeed simpler days.

THE BORN-EINSTEIN LETTERS: CORRESPONDENCE BETWEEN ALBERT EINSTEIN AND MAX AND HEDWIG BORN FROM 1916 TO 1955, WITH COMMENTARIES BY MAX BORN. Translated by Irene Born. Foreword by Bertrand Russell. Introduction by Werner Heisenberg. Walker and Company ($8.50). What a play! What a cast! Wolfgang Pauli, not even mentioned in the billing,

has a brilliant short part in the last act that all but steals the scientific show. The drama, however, is not mainly science. There are some 120 letters, each followed by a comment written by Professor Born between 1965 and 1969. Each bears its own telltale date and its contemporary burden of truth and fancy, of opinion, posturing, irony, advice and jest—whatever went back and forth between Einstein and the Borns as they, their century and ours reeled from the years of Field Marshal Ludendorff and Fritz Haber to those of Joseph McCarthy and Robert Oppenheimer.

There is no mistaking that there are three people on the stage. The role of Hedwig Born—poet, playwright, devout pacifist and generous friend—is wholly engaging. At one point she is turning Einstein's cynical little joke or two about women back on him; at another she is holding before him the ethical vacuousness of a complete determinism. Max Born plays two parts, so to speak: writing recently, he comments on what he and the others thought long ago, creating a play within the play of life. Perhaps this gives the book its special intensity. What happened to this hope and to that plan, what became of the brilliant impoverished assistant? We nearly always learn the outcome; sometimes the very name brings us the news by hindsight. Born wrote: "Young Pauli is very stimulating—I shall never get another assistant as good." Possibly true, and yet his next young assistant was "just as gifted, and also more conscientious." He was Werner Heisenberg.

The depth of the work lies in Einstein. He sounds like Lear—dark, profound, startling, ironic. In 1927 he writes: "The centre of gravity for creative activity is located in different parts of the body in men and women." In 1947 he says that work "helps one bear the craziness of the people who determine the fate of Homo sapiens (so-called) on the grand scale." In 1937, commenting on a professor forced to leave a Muslim university because he is not Islamic, he writes: "It is really rather a comforting thought that in India, too, the all-too-human trait of knavery predominates. After all, it would be just too bad if this were to be the privilege of the proud white race. I believe that all creatures who can have young ones together are very much the same." Also in 1937: "Should I ever recommend a mediocrity to them, even just once, my credit there would be at an end, and I would never again be able to help. It is sad that one is forced to treat human beings like horses where it matters only that they can run and pull, without regard to their qualities as *human beings*."

The concerns of pensions and jobs, of the unfortunate and the promising, fill many of these letters, written so often in war and chaos, in exile or in apprehensive return. The largest theme in point of length of treatment, and the one engaged most technically, is of course the problem of the interpretation of quantum mechanics. Born is the man who suggested the statistical interpretation of the wave function. This interpretation, as Born writes in the late 1960's, is the basis of the common view of the "Copenhagen school, which today lends its name almost everywhere to the line of thinking I originated." Einstein is the adversary, although he wrote in the spring of 1926: "The Heisenberg-Born concepts leave us all breathless.... Instead of dull resignation, there is now a singular tension." Within the year of the birth of the great theory he came to his lifelong doubts: "Quantum mechanics is certainly imposing... but does not really bring us any closer to the secret of the 'old one.' I, at any rate, am convinced that *He* is not playing at dice." Pauli's crisp little essay of 1954, here in the context of a volley of Born and Einstein letters, certainly puts the Copenhagen-Einstein difference in its clearest light. Even it does not have the final word.

It is impossible to avoid comparison between this book, a material witness to the generation of quantum mechanics and of two wars, and another key witness: the recent book by Werner Heisenberg. In the metaphor of the theater, Heisenberg's book is a romantic play—witty and deep, and artfully constructed to display the fate of men and of their ideas. *The Born-Einstein Letters* is closer to today's theater, marked with all haste and imperfection, with real people in real and pressing time, yet with a judicious chorus to throw light (sometimes heat) on events as they were woven into public history and private biography. These are three rare men and women of that old German university world, whose lives "shine with an intense beauty," as Lord Russell, out of the wise generosity of his own ripe age, put it in the foreword.

DEEP OCEANS, edited by Peter J. Herring and Malcolm H. Clarke. Praeger Publishers ($18.50). THE OPEN SEA: ITS NATURAL HISTORY. PART I: THE WORLD OF PLANKTON. PART II: FISH AND FISHERIES, by Sir Alister Hardy. Houghton Mifflin Company ($20). H.M.S. *Challenger*'s four years at sea remains "the longest continual exploration by a scientific vessel." In these books Britannia still rules the science of the waves. *Deep Oceans* is a new overall Anglo-American survey of oceanography (beyond the continental shelf), the majority of whose text and photographs originated with the National Institute of Oceanography, where the editors work. It is a large and attractive book written at a first-reference level, with many graphs and tables as well as full-page color photographs. It aims to convey the look and feel of the modern life of taking ocean data; its longest chapter, titled "Hardware and Software," treats of the sensors towed deep below the ship, the computer on board, the navigational satellite in orbit high overhead. A chapter on ocean physics is particularly well written, its complex material handled by H. Charnock of the University of Southampton with a grace that does not slight the anatomy and physiology of the ocean itself. Most of the book is devoted to the biology of ocean plants and animals and their ecological interactions, and to the exploitation of ocean fisheries. The fisheries chapters, with a final essay on submarine geology, are particularly up to date.

The surface of the open sea is blue, and so are many of the living forms of the neuston: those beings that live in the top few centimeters of the waters. A full-page photograph shows in deepest ultramarine 100-odd little copepods enlarged about 10 times; it is one of the most striking biological pictures in years. The photographer is P. M. David, head of the biology department at the National Institute of Oceanography, and he has furnished more than 40 color pictures showing a glorious living oceanic bestiary from scarlet prawns with lavender egg masses to pale, hairy oceanographers.

The marvelous copepod blue—the result of the combination of the red carotenoid pigment astaxanthin with a protein—is exactly the color of the unboiled lobster, although there it is darkened by other agents. One of its functions is certainly camouflage, yet even powerful predators such as the Portuguese man-of-war have the same pigment. There is a suggestion (accompanied by a graph of absorption v. wavelength) that the pigment reflects the sun's strong near-ultraviolet, which may damage the delicate cells of the animals.

The Open Sea, two normal volumes bound as one, is a different kettle of plankton. It too has many color plates, but they are often watercolor paintings executed by its sole author and chief illustrator, Sir Alister himself. It is per-

sonal and ruminating, a little old-fashioned and romantic. It has few graphs, although there are plenty of maps and sketches. The different flavors of the two books can be compared thus: *Deep Oceans* has real photographs of voracious little squids made with a baited camera in the middle depths. Sir Alister also treats squid life carefully, but he digresses to the giant squid; he presents a photograph of one cast ashore, a mere 20 feet overall, and he cites dark records of 50-foot monsters. *The Open Sea* is not so up to date, particularly in the economics of fisheries. The publisher modestly describes the book as a reissue. (It was completed in 1959 and published in two volumes.) In actuality the first volume has been revised very helpfully, if only here and there, by Sir Alister a decade later.

Pay your money (quite a lot of it) and take your choice. With either book—the personal one or the team effort, color film or watercolor, a background of romance or of biochemistry—you will get a rich and lasting work about the sea.

THE ROLLS-ROYCE STORY, by Lynn Perkins. Rolls-Royce, Inc., Box 564, Paramus, N.J. ($5). Here we review not a book but a poster. Thirty excellent color reproductions, each three inches by four inches, together with their captions fill this single large sheet. They comprise a historical pageant of the magnificent Rolls-Royces, from the 1904 horseless carriage to the latest Silver Shadow. Every car shown here is still on the road; the compiler drove each one of them himself while photographing them. He writes: "I have every confidence that all these machines will be running 1,000 years from now." However that may be, given man's devotion and heaven's corrosion, all the photographs are handsome, presenting some gleaming car against a subdued orange background. The pictures are large enough so that one can just make out that the famous double-R monogram changed from red to black to mourn the death of Sir Henry Royce. Whether the purest automobile type is represented by the 1921 Open Tourer, the aluminum-bodied state car of the Maharaja of Sarila, the 1929 American-built Springfield Speedster Phaeton with its great red wire wheels or the 1907 original Silver Ghost (quite literally silver-plated) individual taste must decide. One picture devoted to a close-up of, say, the Rolls-Royce 40/50 horsepower six-cylinder engine might have been included for those intimate devotees who eagerly lift the bonnet.

INDEX OF ADVERTISERS

AUGUST 1971

BIBLIOGRAPHY

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

CHOLERA

CHOLERA. R. Pollitzer. World Health Organization, Monograph Series No. 43, 1959.

SNOW ON CHOLERA: BEING A REPRINT OF TWO PAPERS BY JOHN SNOW, M.D. TOGETHER WITH A BIOGRAPHICAL MEMOIR BY B. W. RICHARDSON, M.D., AND AN INTRODUCTION BY WADE HAMPTON FROST, M.D. Hafner Publishing Company, 1965.

THE ROLE OF CHOLERA EXOTOXIN IN THE STUDY OF INTESTINAL WATER AND ELECTROLYTE TRANSPORT. William B. Greenough III, Charles C. J. Carpenter, Theodore M. Bayless and Thomas R. Hendrix in *Progress in Gastroenterology: Vol. II*, edited by George B. Jerzy Glass. Grune & Stratton, 1970.

A BYZANTINE TRADING VENTURE

THE RHODIAN SEA-LAW. Edited by Walter Ashburner. Oxford University Press, 1909.

BYZANTINE TRADE AND INDUSTRY. Steven Runciman in *The Cambridge Economic History of Europe, Vol. II: Trade and Industry in the Middle Ages*, edited by M. Postan and E. E. Rich. Cambridge University Press, 1952.

REPLICAS OF IRON TOOLS FROM A BYZANTINE SHIPWRECK. Michael L. Katzev and Frederick H. van Doorninck, Jr., in *Studies in Conservation*, Vol. 2, No. 3, pages 133–141; August, 1966.

A FOURTH-CENTURY SHIPWRECK AT YASSI ADA. George F. Bass and Frederick H. van Doorninck, Jr., in *American Journal of Archaeology*, Vol. 75, No. 1, pages 27–37; January, 1971.

HIGH-LYSINE CORN

PROCEEDINGS OF THE HIGH LYSINE CORN CONFERENCE. Edited by Edwin T. Mertz and Oliver E. Nelson. Corn Refiners Association, Washington, D.C., 1966.

WORLD PROTEIN RESOURCES: ADVANCES IN CHEMISTRY SERIES, 57. American Chemical Society, 1966.

GENETIC MODIFICATION OF PROTEIN QUALITY IN PLANTS. Oliver E. Nelson in *Advances in Agronomy*, Vol. 21, pages 171–194; 1969.

SYMBIOSIS AND EVOLUTION

THE PLASTIDS: THEIR CHEMISTRY, STRUCTURE, GROWTH AND INHERITANCE. John T. O. Kirk and Richard A. E. Tilney-Bassett. W. H. Freeman and Company, 1967.

THE BIOGENESIS OF MITOCHONDRIA. D. B. Roodyn and D. Wilkie. Methuen & Co Ltd, 1968.

THE MICROBIAL WORLD. Roger Y. Stanier, Michael Doudoroff and Edward A. Adelberg. Prentice-Hall, Inc., 1970.

ORIGIN OF EUKARYOTIC CELLS. Lynn Margulis. Yale University Press, 1970.

THE OLDEST FOSSILS. Elso S. Barghoorn in *Scientific American*, Vol. 224, No. 5, pages 30–42; May, 1971.

THE MAGNETISM OF THE MOON

MAGNETISM AND THE COSMOS. Edited by W. R. Hindmarsh, F. J. Lowes, P. H. Roberts and S. K. Runcorn. Oliver & Boyd, 1967.

THE MOON: OUR NEAREST CELESTIAL NEIGHBOUR. Zdeněk Kopal. D. Reidel Publishing Company, 1969.

LUNAR PORTABLE MAGNETOMETER EXPERIMENT. P. Dyal, C. W. Parkin, C. P. Sonett, R. L. DuBois and G. Simmons in *Apollo 14 Preliminary Science Report*. NASA SP–272. U.S. Government Printing Office, July, 1971.

THE APOLLO 12 MAGNETOMETER EXPERIMENT: INTERNAL LUNAR PROPERTIES FROM TRANSIENT AND STEADY MAGNETIC FIELD MEASUREMENTS. P. Dyal and C. W. Parkin in *Proceedings of the Second Lunar Science Conference: Houston, Texas, January 11–14, 1971*. The M.I.T. Press, in press.

FLIGHT ORIENTATION IN LOCUSTS

THE ORIENTATION OF ANIMALS: KINESES, TAXES AND COMPASS REACTIONS. Gottfried S. Fraenkel and Donald L. Gunn. Dover Publications Inc., 1961.

THE CENTRAL NERVOUS CONTROL OF FLIGHT IN A LOCUST. Donald M. Wilson in *The Journal of Experimental Biology*, Vol. 38, No. 2, pages 471–490; June, 1961.

REFLEXES AND MOTOR INTEGRATION: SHERRINGTON'S CONCEPT OF INTEGRATIVE ACTION. Judith P. Swazey. Harvard University Press, 1969.

YAW-CORRECTING POSTURAL CHANGES IN LOCUSTS. Jeffrey M. Camhi in *The Journal of Experimental Biology*, Vol. 52, No. 3, pages 519–531; June, 1970.

THE CONTROL OF SHORT-TERM MEMORY

HUMAN MEMORY: A PROPOSED SYSTEM AND ITS CONTROL PROCESSES. R. C. Atkinson and R. M. Shiffrin in *Advances in the Psychology of Learning and Motivation Research and Theory: Vol. II*, edited by K. W. Spence and J. T. Spence. Academic Press, 1968.

MEMORY SEARCH. Richard M. Shiffrin in *Models of Human Memory*, edited by Donald A. Norman. Academic Press, 1970.

FORGETTING: TRACE EROSION OR RETRIEVAL FAILURE. Richard M. Shiffrin in *Science*, Vol. 168, No. 3939, pages 1601–1603; June 26, 1970.

HUMAN MEMORY AND THE CONCEPT OF REINFORCEMENT. R. C. Atkinson and T. D. Wickens in *The Nature of Reinforcement*, edited by R. Glaser. Charles E. Merrill Books, Inc., 1971.

AN ANALYSIS OF REHEARSAL PROCESSES IN FREE RECALL. D. Rundus in *Journal of Experimental Psychology*, in press.

NEW MODELS OF THE REAL-NUMBER LINE

NON-STANDARD ANALYSIS. Abraham Robinson. North-Holland Publishing Company, 1966.

A PROOF OF THE INDEPENDENCE OF THE CONTINUUM HYPOTHESIS. Dana Scott in *Mathematical Systems Theory*, Vol. 1, No. 2, pages 89–111; Summer, 1967.

NON-CANTORIAN SET THEORY. Paul J. Cohen and Reuben Hersh in *Scientific American*, Vol. 217, No. 6, pages 104–116; December, 1967.

MATHEMATICAL GAMES

THE THEORY OF GAMBLING AND STATISTICAL LOGIC. Richard A. Epstein. Pages 359–363. Academic Press, 1967.

THE GAME OF NOUGHTS AND CROSSES. Fred Schuh in *The Master Book of Mathematical Recreations*. Dover Publications, Inc., 1968.

THE AMATEUR SCIENTIST

INTRODUCTORY NUCLEAR PHYSICS. David Halliday. John Wiley & Sons, Inc., 1955.

AN INTRODUCTION TO RADIATION COUNTERS AND DETECTORS. C. C. H. Washtell. Philosophical Library Inc., 1958.

WHY VEGA IS PRICED A LITTLE HIGHER THAN OTHER LITTLE CARS.

We admit that Vega isn't the most inexpensive little car you can buy.

Nonetheless, we maintain that, dollar for dollar, Vega is the biggest bargain you can get in little cars.

And we have evidence.

Now you see it.

There are other little cars that came out this year, just like Vega.

Only unlike Vega, they have engines that already existed in other cars. Not Vega. We designed an engine (1) from scratch. Overhead cam. Aluminum block. It'll give you plenty of power. And, in our highway tests, we've been getting about 25 mpg with the standard engine and transmission.

There are other obvious differences between Vega and other little cars, too. Like our wheels and tires (2). They're a little wider than the wheels and tires on most little cars. A little thing? We think anything we can do to make Vega ride and handle better is a big thing indeed.

Another not-so-little thing. The back seat (3). You may never have to use the back seat yourself. But somebody will. And when they do, they'll discover that Vega's back seat is a foam-cushioned seat, complete with frame. Now we realize that may not sound like any big deal, but when you're out comparing little cars, try out the back seats.

The difference is astounding.

Now you don't.

Some of the differences between Vega and most other little cars just can't be seen with the naked eye.

But they're real all the same.

Like protective side-guard door beams (4), for instance. No other little car has them. Vega does.

And front disc brakes (5). Some little cars we know don't have them. But Vega does—nice big 10-inchers. Because we want you to have good stopping power.

And Vega also has a power ventilation system (6). And an acoustically engineered double-panel roof (7). And an electric fuel pump (8). And coil springs at each wheel (9).

And, well, more.

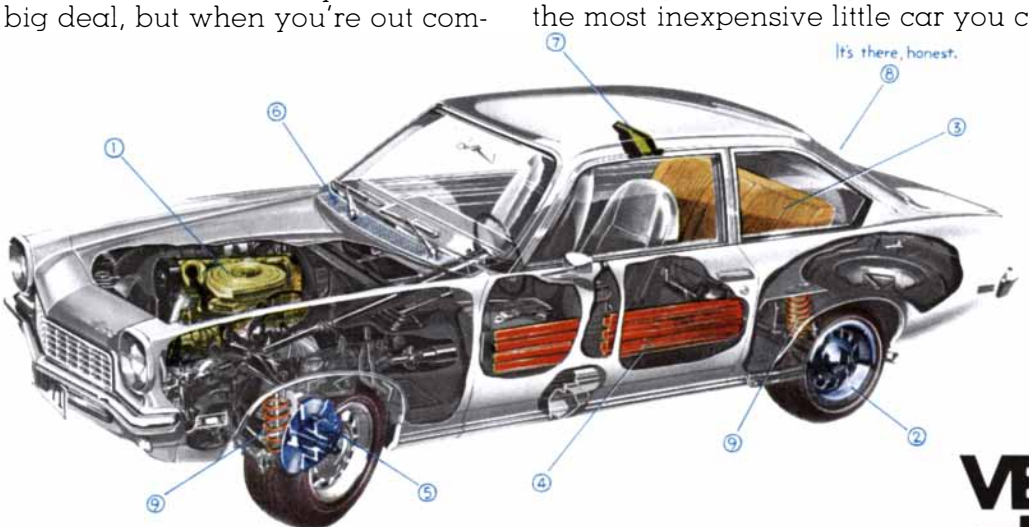
The moral.

We built Vega on the premise that a little car should be just as much car as any big car, only smaller.

And Vega is.

If you still doubt us, compare. Come on in and sit in our seats. Take us for a ride. Read our catalog. Ask us questions.

Vega. In the long run, it just might be the most inexpensive little car you can buy.

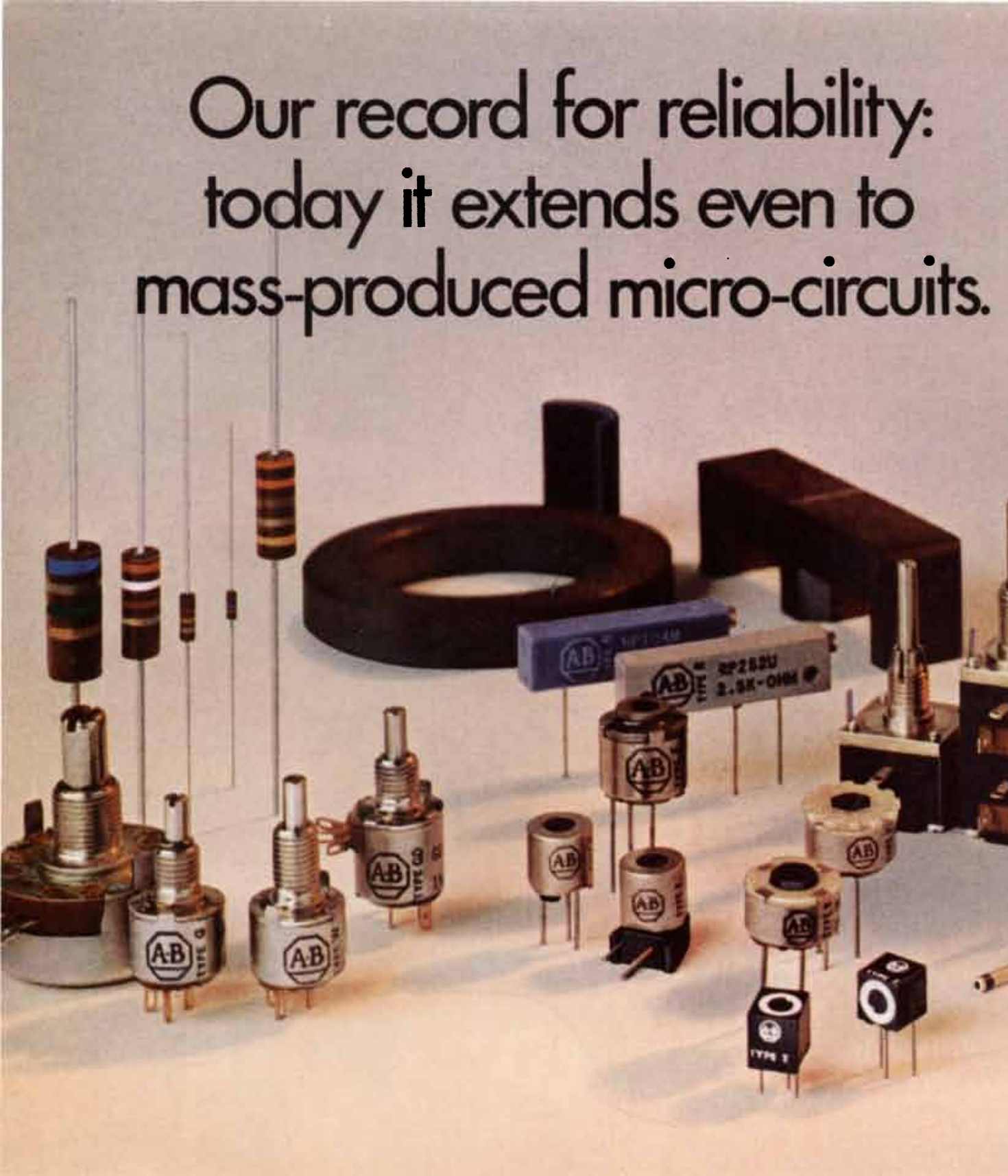


GM
MARK OF EXCELLENCE

VEGA
CHEVROLET

Buckling your seat and shoulder belts is an idea you can live with.

Our record for reliability: today it extends even to mass-produced micro-circuits.



You know our resistors, billions produced to the highes quality standards. Now we've added to this tradition. Pr thick and thin film networks. Tiny RFI filters to conquer pollution. Miniaturized pots and trimmers. Hybrids, opt electronic devices, even large scale integrated circuits

*LSI devices produced by an Allen-Bradley affiliate, MOS Technol